Technical Physics, Vol. 47, No. 9, 2002, pp. 1071–1082. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 1–12.

THEORETICAL AND MATHEMATICAL PHYSICS

Computer Approach to the Determination of Electrostatic Fields and Capacitive Coefficients for 3D Multielement Sets of Screens

S. I. Safronov and R. P. Tarasov

Research Institute of Pulsed Technology, Moscow, 115304 Russia

e-mail: doi@ript.in.ru

Received October 19, 2001; in final form, January 24, 2002

Abstract—The structural components of a computer handbook on electrostatic fields and capacitive coefficients of 3D multielement sets of screens are described. The computer approach suggested is based on boundary integral equations from the 3D theory of harmonic potentials and fast algorithms of the finite-group method that are realized on a Pentium-Pro 200 MHz 128-MB RAM PC. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The creation of radio engineering devices in many cases requires the determination of electrostatic fields and capacitive coefficients for multielement sets of screens. For a given set of screens of specific configuration, the solution of these basic problems of electrostatics in the 3D case is known to be associated with substantial difficulties. This, as well as a variety of electrostatic systems that should be analyzed in designing radio engineering devices, cause researchers to discard 3D models in favor of qualitative 2D ones, which are less accurate. However, many multielement systems of a rather general form can be designed by using an appropriate computer handbook involving fast algorithms for numerically solving boundary integral equation of the first kind from the 3D theory of harmonic potentials.

A computer handbook of electrostatic fields and capacitive coefficients must allow researchers to design systems close to or even coincident with a desired system from a limited yet representative set of basic components. It seems to be natural that the starting set of basic components be as follows: flat symmetric screens (triangles, rectangles, and circles), surfaces of revolution, parallelepipeds, and regular polyhedra. It is assumed that all can have holes of a specific shape and that the starting set of components can be supplemented according to the design practice. The handbook must provide a designer with electrostatic fields, as well as self- and mutual capacitances, for a given layout of basic components incorporated into a 3D set of screens.

Below, we will demonstrate that the computer handbook suggested can be based on fast algorithms used in the finite-group method (FGM). With this method, the computer simulation of 3D electron–optic systems [1, 2] and the scattering of the electrostatic field of a TEM chamber by local inhomogeneities [3] has been performed.

STARTING RELATIONSHIPS IN THE HARMONIC POTENTIAL METHOD

Consider a set $\{S_i\}$ of N conducting screens S_i under the assumption that $\{S_i\}$ produces a multiply connected surface S in a 3D space R^3,

$$S = \bigcup_{i=1}^{N} S_i, \quad S_i \cap S_j = \emptyset, \quad i \neq j, \tag{1}$$

and that each of the screens S_j is under a given potential V_j, $j = 1, 2, \ldots, N$. If $E = \{E_i\}(i = 1, 2, 3)$ ($i = 1, 2, 3$) is the vector of an electric field induced by the set of screens S, then $E_i(x) = -\partial\varphi(x)/\partial x_i$, where $x = \{x_i\}$ are the Cartesian coordinates of a point x in R^3 and $\varphi(x)$ is the potential of the field E that satisfies the Laplace equation with the Dirichlet conditions on S:

$$\begin{gathered}\Delta\varphi(x) = 0, \quad \Delta \equiv \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2 + \partial^2/\partial x_3^2, \\ x \in R^3 \backslash S, \\ \varphi(x) = \sum_{i=1}^{N} \chi_{S_i}(x) V_i, \quad x \in S.\end{gathered} \tag{2}$$

Here, $\chi_{S_i}(x)$ is the characteristic function of a set of points on the surface S_i.

A solution to boundary-value problem (2) can be represented as the potential of a layer,

$$\varphi(x) = \int_S \frac{u(x_1)}{|x - x_1|} d\mu(x_1), \quad x_1 \in R^3 \backslash S,$$

1063-7842/02/4709-1071$22.00

with a density $u(x)$ ($x \in S$) satisfying the boundary integral equation of first kind

$$[Au](x) = f(x), \quad [Au](x) = \int_S \frac{u(x_1)}{|x - x_1|} d\mu(x_1),$$
$$f(x) = \sum_{i=1}^{N} \chi_{S_i}(x) V_i, \tag{3}$$

where $d\mu(x)$ is the contraction of a Euclidean volume element on the surface S and $|x - x_1|$ is the Euclidean spacing between points x and x_1.

Hence, the electrostatic field of the system (set) S of the screens will be defined if a solution to the boundary integral equation (3) is constructed. Next, if Q_i is the charge of a screen S_i,

$$Q_i = \int_S \chi_{S_i}(x) u(x) d\mu(x), \tag{4}$$

and Q and V are considered as column matrices, $Q = (Q_1, Q_2, \ldots, Q_N)$ and $V = (V_1, V_2, \ldots, V_N)$, one can write [4]

$$Q = cV, \tag{5}$$

where $c = \|c_{mn}\|$ ($m, n = 1, 2, \ldots, N$) is the matrix of the capacitive coefficients of the system S.

Putting $V_q = 1$ and $V_i = 0$ at $i \neq q$, we have from (4) and (5)

$$c_{iq} = \int_S \chi_{S_i}(x) u(x) d\mu(x), \quad i = 1, 2, \ldots, N. \tag{6}$$

Hence, the matrix $\|c_{mn}\|$ of the capacitive coefficients can be constructed by solving the family of boundary-value problems.

The associated solution is given by

$$\Delta\varphi^{(q)}(x) = 0, \quad x \in R^3 \backslash S,$$
$$\varphi^{(q)}(x) = \sum_{i=1}^{N} \chi_{S_i}(x) \delta_{iq}, \quad x \in S, \quad q = 1, 2, \ldots, N, \tag{7}$$

where δ_{iq} is the Kronecker symbol. It has the form

$$\delta_{iq} = \begin{cases} 1, & i = q \\ 0, & i \neq q \end{cases}$$

provided that solution (7) is derived with boundary integral equations like

$$Au^{(q)}(x) = \sum_{i=1}^{N} \chi_{S_i}(x) \delta_{iq}, \quad x \in S, \quad q = 1, 2, \ldots, N. \tag{8}$$

Indeed, from (6)–(8) for the elements of the matrix $\|c_{mn}\|$, we can find

$$c_{iq} = \int_S \chi_{S_i}(x) u^{(q)}(x) d\mu(x), \quad i, q = 1, 2, \ldots, N. \tag{9}$$

Furthermore, since

$$\varphi(x) = \sum_{q=1}^{N} V_q \varphi^{(q)}(x), \quad x \in R^3,$$

the solution of the two basic problems of electrostatics (the determination of the electrostatic field for given potentials and of the capacitive coefficient matrix) for the 3D system S of the screens is reduced to a solution of family (8) of boundary integral equations.

If the surface S is simply connected and has a simple shape, Eq. (3) can be solved numerically and highly accurately by the boundary-element method. If the surface S is multiply connected and has a complex shape, the standard boundary-element method generates a need for inverting dense matrices of a very high dimension (several tens of thousands or more), for which a fast yet accurate inversion procedure is difficult to devise within conventional approaches.

For electrostatic systems of more or less general type [1, 2], fast and highly accurate algorithms for numerically solving Eq. (3) can be constructed with the FGM, which exploits local symmetries in subsystems into which an initial system is divided.

BASIC IMPLEMENTATIONS OF THE FINITE-GROUP METHOD IN ELECTROSTATICS PROBLEMS

The FGM relies on the invariance of the Laplace, Helmholtz, and Maxwell equations under the motion of a Euclidean space [5, 6] and uses the structural relations in Fourier geometrical analysis of invariant boundary operators on a surface with a disconnected group [7–9]. If a scattering body possesses an N-fold group of symmetries, one can construct a fast algorithm that reduces the number of operations by a factor of $\sim N^2$, representing the boundary equations as an operator convolution on the group of the body's symmetries [8]. An asymmetric problem can be included, under certain conditions, in the class of symmetric problems by using the gluing approach. In this case, the boundary surface S of an initial boundary-value problem without symmetry is embedded in the surface $\bar{S}$ ($S \subset \bar{S}$) of a problem with a finite group of symmetries. Then, the initial boundary-value problem can be stated in the form of two subproblems, one having a finite group of symmetries and the other admitting a relatively simple numerical solution [10, 11]. Note also that the partition of the initial surface S (without symmetry) into constituents that have a finite group of symmetries and the subsequent iterative sewing based on triangular schemes allow the extension of the FGM for a wide class of complicated 3D

boundary-value problems of great applied interest [12–14].

(1) Let

$$Au = f, \quad [A\Psi](x) = \int_{x, x_1 \in S} A(x, x_1)\Psi(x_1)d\mu(x_1), \tag{10}$$

be a boundary integral equation of the first kind for the Dirichlet problem of the Laplace equation and let representation (1) be valid for the surface S. Applying (1), we relate the function $\varphi(x)$ (defined on the multiply connected surface S) to some function $\varphi_i(x)$ taking (at each fixed i, $i = 1, 2, \ldots, N$) a value in the space of functions defined on S_i,

$$\varphi(x) \longrightarrow \varphi_i(x) = \chi_{S_i}(x)\varphi(x), \tag{11}$$

and the boundary operator A on the multiply connected surface S, to the operator matrix $\|A_{ij}\|$,

$$A \longrightarrow \|A_{ij}\|, \quad A_{ij} = \chi_{S_i}(x)A\chi_{S_j}(x), \quad x \in S, \quad i, j = 1, 2, \ldots, N. \tag{12}$$

At fixed i and j, the operator A_{ij} maps from the space of functions defined on S_j into the space of functions defined on S_i. Then, in view of representations (11) and (12), boundary integral equation (10) can be written in the form

$$\sum_{j=1}^{N} A_{ij}u_j(x) = f_i(x), \quad x \in S_i, \quad i = 1, 2, \ldots, N. \tag{13}$$

A solution to Eq. (13) will be constructed by iterations with the use of a matrix triangular scheme:

$$A_{ii}u_i^{(k+1)} = (1-\omega)A_{ii}u_i^{(k)} + \omega\left[f_i - \sum_{j<i} A^{ij}u_j^{(k+1)} - \sum_{j>1} A_{ij}u_j^{(k)}\right], \tag{14}$$

$$i = 1, 2, \ldots, N; \quad k = 0, 1, 2, \ldots; \quad \omega \in (0, 2).$$

Note that for a specific parameter ω, $\omega \in (0, 2)$, iteration procedure (14) offers a high rate of convergence for problems of a certain class (in electrostatic problems, for $\omega = 1$) and admits the representation

$$u_i^{(k+1)} = (1-\omega)u_i^{(k)} + \omega A_{ii}^{-1}\left[f_i - \sum_{j<i} A_{ij}u_j^{(k+1)} - \sum_{j>1} A_{ij}u_j^{(k)}\right], \tag{15}$$

$$i = 1, 2, \ldots, N; \quad k = 0, 1, 2, \ldots,$$

which is suitable for computer realization.

In this case, computational resources are spent basically on the construction of the operators A_{ii}^{-1} reciprocal to A_{ii} ($i = 1, 2, \ldots, N$). The reciprocal operators are constructed once before iterations.

(2) The multiply connected surface S comprises basic elements S_i ($i = 1, 2, \ldots, N$), each being described (initially or after gluing) by a finite group of symmetries. This symmetry property can be realized directly in a computational algorithm for inverting operators A_{ii} at any fixed i ($i = 1, 2, \ldots, N$). Introducing the designation

$$F_i^{(k)} = \omega\left[f_i - \sum_{j<1} A_{ij}u_j^{(k+1)} - \sum_{j>i} A_{ij}u_j^{(k)}\right],$$

$$i = 1, 2, \ldots, N; \quad k = 0, 1, 2, \ldots,$$

we rewrite the ith equation of iteration process (15) as

$$u_i^{(k+1)} = (1-\omega)u_i^{(k)} + z_i^{(k)},$$

where $z_i^{(k)}$ is a solution to the equation

$$[A_{ii}z_i^{(k)}](x) = F_i^{(k)}(x), \quad x \in S_i. \tag{16}$$

Let $\{\tau_M\}$ be a group of symmetries of an element S_i on the surface S (M is the order of the group). Then, S_i can be represented as

$$S_i = \bigcup_{\tau_m \in \{\tau_M\}} \tau_m s_0, \quad \mathrm{int}(\tau_m s_0 \cap \tau_n s_0) = 0, \quad m \neq n, \tag{17}$$

where s_0 is the fundamental region of the group $\{\tau_M\}$ disconnected on S.

Let us introduce the invariant partition of unity that corresponds to representation (17):

$$I = \sum_{\tau_m \in \{\tau_M\}} \chi_{s_0}(\tau_m^{-1}\tau_n x) \ \forall \tau_n \in \{\tau_M\}, \tag{18}$$

where $\chi_{s_0}(\cdot)$ is the characteristic function of the fundamental region s_0 and the equality is fulfilled up to the set of measure zero.

Using partition of unity (18), we relate each of the functions $\varphi_i(x)$, $x \in S_i$, to a function $\hat{\varphi}(\tau_m)$, $\tau_m \in \{\tau_M\}$, from the group $\{\tau_M\}$. The latter function has values in the space of functions $L(s_0)$ that are defined on s_0,

$$\varphi_i(x) \longrightarrow \hat{\varphi}(\tau_m) = \chi_{s_0}(\tau_m^{-1}x)\varphi_i(x).$$

The operator A_{ii} is related to an operator function $\hat{\alpha}(\tau_m, \tau_n)$ with values in the set of operators $\{\tau_M\}$ acting in $L(s_0)$:

$$A_{ii} \longrightarrow \hat{\alpha}(\tau_m, \tau_n) = \chi_{s_0}(\tau_m^{-1}x)A_{ii}\chi_{s_0}(\tau_n^{-1}x), \quad \tau_m, \tau_n \in \{\tau_M\}.$$

The operator of the first-kind integral equation of the boundary-value Dirichlet problem for the Laplace

equation is invariant under symmetry transformation of the boundary surface. In addition, we have

$$A_{ii} = \chi_{S_i}(x) A \chi_{S_i}(x).$$

From this directly follows the invariance of the operator A_{ii} under the actions of the group $\{\tau_M\}$; therefore, the associated operator function $\hat{\alpha}(\tau_m, \tau_n)$ is invariant under left shifts on the group $\{\tau_M\}$:

$$\hat{\alpha}(\tau_l\tau_m, \tau_l\tau_n) = \hat{\alpha}(\tau_m, \tau_n), \quad \forall \tau_l \in \{\tau_M\}$$

and we can write the equality

$$\sum_{\tau_n \in \{\tau_M\}} \hat{\alpha}(\tau_m, \tau_n)\hat{\varphi}(\tau_n) = \sum_{\tau_n \in \{\tau_M\}} \hat{\alpha}(\tau_n^{-1}\tau_m)\hat{\varphi}(\tau_n),$$

$$\hat{\alpha}(\tau_m) \equiv \hat{\alpha}(\tau_m, e),$$

where e is unity of the group $\{\tau_M\}$.

Therefore, for an element S_i with the group of symmetries $\{\tau_M\}$ of the surface S, Eq. (16) takes the form of an equation for convolution on the group [8]:

$$\sum_{\tau_n \in \{\tau_M\}} \hat{\alpha}(\tau_n^{-1}\tau_m)\hat{z}(\tau_n) = \hat{F}(\tau_m), \quad \tau_m \in \{\tau_M\}. \quad (19)$$

Let U_λ, $\lambda \in \{\lambda\}$, be irreducible representations of the group $\{\tau_M\}$ and $d\lambda$ be the degree of the representation U_λ. We introduce operator coefficients of the Fourier functions $\hat{\alpha}(\tau_n)$, $\hat{z}(\tau_n)$, and $\hat{F}(\tau_n)$ on the group $\{\tau_M\}$:

$$\tilde{\alpha}(\alpha) = \sum_{\tau_n \in \{\tau_M\}} \hat{\alpha}(\tau_n) U_\lambda(\tau_n^{-1}), \quad \lambda \in \{\lambda\},$$

$$\tilde{z}(\lambda) = \sum_{\tau_n \in \{\tau_M\}} \hat{z}(\tau_n) U_\lambda(\tau_n^{-1}),$$

$$\tilde{F}(\lambda) = \sum_{\tau_n \in \{\tau_M\}} \hat{F}(\tau_n) U_\lambda(\tau_n^{-1}), \quad \lambda \in \{\lambda\}.$$

By virtue of Eq. (19), we can write

$$\tilde{\alpha}(\lambda) z(\lambda) = \tilde{F}(\lambda), \quad \lambda \in \{\lambda\}. \quad (20)$$

If $z(\lambda)$ is a solution to Eq. (20), the function $\hat{z}(\tau_n)$, $\tau_n \in \{\tau_M\}$, is calculated from the relationship

$$\hat{z}(\tau_n) = \sum_{\lambda \in \{\lambda\}} d_\lambda tr(\hat{z}(\lambda) U_\lambda(\tau_n)), \quad \tau_n \in \{\tau_M\},$$

where $tr(\cdot)$ means trace.

The construction of the operators A_{ii}^{-1} by calculating the operators $\tilde{\alpha}^{-1}(\lambda)$ reciprocal to $\tilde{\alpha}(\lambda)$, $\lambda \in \{\lambda\}$, is $\sim M^2$ times more optimal (in terms of the number of operations) than direct inversion of the operator A_{ii} with the standard procedures.

Now, let an element S_i of the surface S have an associated symmetry only after the gluing procedure (basic screens with holes); that is, let there exist a surface S_i with the group of symmetries $\{\tau_M\}$ such that

$$\bar{S}_i = \bigcup_{\tau_m \in \{\tau_M\}} \tau_m \bar{s}_0, \quad \text{int}(\tau_m \bar{s}_0 \cap \tau_n \bar{s}_0) = 0, \quad m \neq n.$$

Here, $\bar{s}_0$ is the fundamental region of the group $\{\tau_M\}$ such that

$$S_i \subset S_i, \quad S_i^{(0)} = S_i \backslash S_i$$

and the measure of the set $S_i^{(0)}$ is less than that of the set S_i:

$$\mu(S_i^{(0)}) < \mu(S_i). \quad (21)$$

In this case, we pass from Eq. (16) to the problem

$$\chi_{S_i}(x) B \nu(x) = \chi_{S_i}(x) \eta(x), \quad x \in S_i,$$

$$\chi_{S_i^{(0)}}(x) \nu(x) = 0, \quad x \in S_i^{(0)},$$

where

$$[B\Psi](x) = \int A(x, x_1)\Psi(x_1) d\mu(x_1),$$

$$x, x_1 \in \bar{S}_i, \quad \chi_{S_i}(x) B \chi_{S_i}(x) = A_{ii}$$

and

$$\eta(x) = \chi_{S_i}(x)\eta(x) + \chi_{S_i^{(0)}}(x)\eta(x), \quad x \in \bar{S}_i,$$

$$\chi_{S_i}(x)\eta(x) = F_i^{(k)}(x), \quad x \in S_i.$$

The transition from (16) to (22) is associated with two main factors: the trace of the solution to (22) on S_i is the solution to (16),

$$\chi_{S_i}(x) \nu(x) = z_i^{(k)}(x), \quad x \in S_i,$$

and the operator B is invariant under the action of the finite group $\{\tau_M\}$. The latter fact makes it possible [10, 11] to construct a solution algorithm for (22) that is optimal in terms of the number of operations. Subject to (21), this algorithm shrinks the number of operations for inverting the operator A_{ii} by a factor of $\sim M^2$ in comparison with the direct construction of the operator A_{ii}^{-1} reciprocal to A_{ii}. A more detailed numerical solution of boundary-value problems as applied to the Laplace equation within the FGM was considered in [15].

COMPUTER DETERMINATION OF ELECTROSTATIC FIELDS AND CAPACITIVE COEFFICIENTS FOR 3D SETS OF SCREENS

Based on the fast algorithms within the finite-group method, we created a suite for determining electrostatic

fields and capacitive coefficients for 3D sets of screens composed of the basic components listed in the Introduction. For a system to be designed, basic components and their layout are selected in an arbitrary way. The maximal number of the screens depends only on the computer power and the analytical accuracy.

(1) Consider a system of two spheres having radii R (sphere S_R) and $r < R$ (sphere S_r). We assume that the center of S_r is shifted relative to that of S_R by a distance $H = h(R - r)$ $(0 \le h \le 1)$. If $h = 0$, the centers of S_R and S_r coincide (spherical capacitor), while at $h = 1$, the spheres touch each other at a single point. According to [4], the normalized dimensionless capacitance $C(h)$ of such a system is given by (for $h \in [0, 1]$)

$$C(h) = \frac{Q_r(h)}{4\pi\varepsilon r(V_r - V_R)},$$

where V_R and V_r are the potentials of the spheres S_R and S_r, respectively, $Q_r(h)$ is the charge on the sphere S_r, and ε is the permittivity of the medium.

For $R/r = 2$ and several $h \in [0, 1]$, the calculated values of $C(h)$ are listed in Table 1.

The surface S, $S = S_R \cup S_r$, was covered by 2×10^3 boundary components. The exact value of $C(h)$ at $h = 0$ is found from the formula given in [4]: $C_0 = R/(R - r) = 2$. Thus, our value of $C(0)$ is accurate within 0.1%. Note that in order to achieve computer accuracy, the number of boundary elements covering S must be increased to $\sim 5 \times 10^3$.

Equipotential lines in the plane containing the centers of both spheres are shown for $V_R = 0$, $V_r = 1V$, $R = 2m$, $r = 1m$, and $h = 0$ (Fig. 1a) and 0.8 (Fig. 1b).

(2) Shushkevich [16] analyzed the axisymmetric electrostatic problem for a thin nonclosed spherical shell s and a torus T of small radius r and large radius R. The shell s is placed on a sphere S_1 of radius d. Consider another system comprising the same shell s and a torus T_1 with a rectangular cross section. The axial sections of both systems are depicted in Fig. 2a.

Table 1

h	$C(h)$
0	1.998
0.2	2.0202
0.4	2.1022
0.6	2.2734
0.8	2.6739

If $c = \|c_{ij}\|$ $(i, j = 1, 2)$ is the matrix of capacitive coefficients for a double-electrode system, the capacitance C of the system is calculated from the formula

$$C = \frac{c_{11}c_{22} - c_{12}c_{21}}{c_{11} + c_{12} + c_{21} + c_{22}}.$$

The normalized capacitances $\tilde{C} = C/(4\pi\varepsilon d)$ for $R/r = 2$, two ratios R/d, and several apex angles Θ of the nonclosed thin spherical shell s are listed in Table 2. The values of $\tilde{C}$ for the first ($S^{(1)} = s \cup T$) and second ($S^{(2)} = s \cup T_1$) systems are given in the columns T and T_1, respectively.

It is clear that $\tilde{C}$ for the system $S^{(1)}$ and $\tilde{C}$ obtained in [16] differ by no more than 1.8%. At the same time, it is obvious that the capacitances of the systems $S^{(1)}$ and $S^{(2)}$ as functions of the apex angle Θ and ratio R/r depend mostly on the mutual arrangement and size of the screens and to a lesser extent on their shape. This fact gives certain grounds for replacing actual sets of screens by sets of approximating basic components.

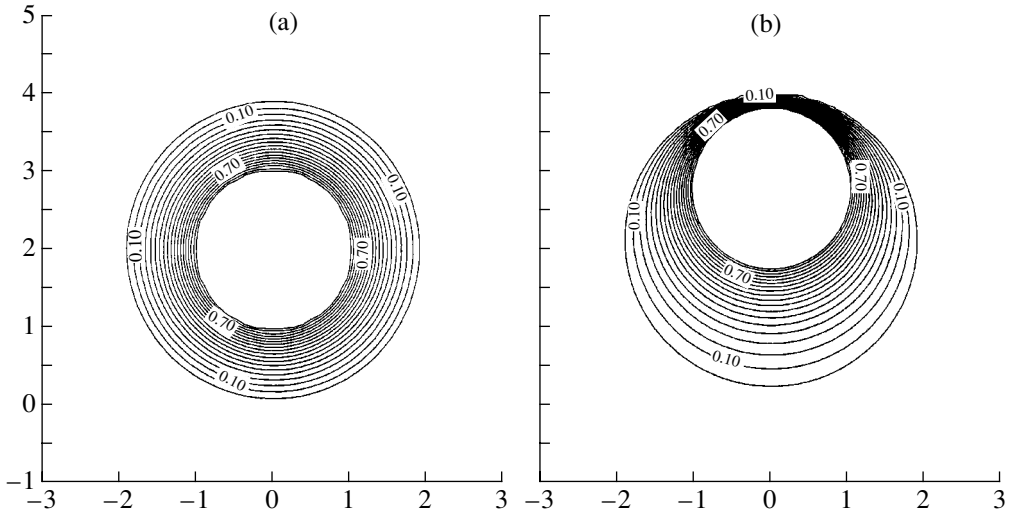


Fig. 1.

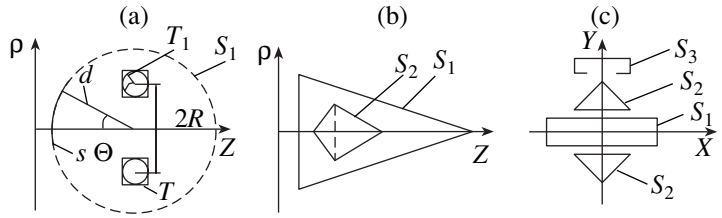


Fig. 2.

Here, some comments should be made. Screens of each of the three sets considered above produce a multiply connected surface of revolution. In this case, a computational algorithm for solving Eq. (10) optimized in terms of the number of operations is constructed according to the technique described in [17]. If the right-hand side of (10) is invariant under shift on group of rotation C_∞ (Eq. (3) and family of equations (8) are examples), Eq. (10) can be rewritten as a 1D equation over the generator of the surface of revolution. Methods to attack such 1D equations have been suggested in [18]. At present, axisymmetric problems of electrostatics are solved by the FGM with a very high accuracy and with minimal machine costs for almost any arbitrary surface of revolution.

Consider how the capacitance of the system $S^{(2)}$ with $\Theta = 90°$, $R/r = 2$, and $R/d = 0.5$ varies with the angle α made by the axes of the hemisphere s and torus T_1. For the normalized capacitance $\tilde{C}(\alpha)$, we have

$$\tilde{C}(0°) = 1.548, \quad \tilde{C}(45°) = 1.517,$$
$$\tilde{C}(90°) = 1.493$$

and the matrix $c(\alpha)$ of capacitive coefficients takes the form

$$c(0°) = \begin{bmatrix} 8.10071 & -5.52312 \\ -5.52312 & 6.42901 \end{bmatrix},$$

Table 2

Θ, deg	Capacitance			
	$R/d = 0.5$		$R/d = 0.1$	
	T	T_1	T	T_1
10	0.111	0.117	0.653	0.678
30	0.334	0.359	0.108	0.115
60	0.754	0.837	0.128	0.138
90	1.291	1.548	0.135	0.146
120	1.592	1.977	0.138	0.150
150	1.612	2.042	0.139	0.151
170	1.615	2.048	0.139	0.151

$$c(45°) = \begin{bmatrix} 8.06476 & -5.40814 \\ -5.40814 & 6.28437 \end{bmatrix},$$

$$c(90°) = \begin{bmatrix} 7.99781 & -5.31049 \\ -5.31049 & 6.19082 \end{bmatrix}.$$

Note that all elements c_{ij} ($i, j = 1, 2$) of the matrices $c(\alpha)$ are found by calculation; therefore, the relationship $c_{12} = c_{21}$ serves as a criterion of calculation accuracy.

Equipotential lines for the system $S^{(2)}$ with $V_s = 0$ and $V_{T_1} = 1V$, where V_s is the potential of the hemisphere and V_{T_1} is that of the torus, in the plane containing the axes of revolution of the hemisphere and torus are depicted in Fig. 3 for $\alpha = 0°$ (a), 45° (b), and 90° (c). Figure 3d ($\alpha = 90°$) also shows equipotential lines in the plane containing the axis of revolution of the hemisphere and the orthogonal axis of revolution of the torus T_1. The equipotentiality of the screens is clearly demonstrated (the field near the screen copies its shape). From Fig. 3d, it also follows that the decay rate of the disturbance is high (the shape of the screen rapidly diffuses away from it).

(3) Consider several two-component systems S comprising a canonical body of revolution (sphere, cylinder, or cone) and a body made up of two cones with a mutual base and oppositely directed heights. The capacitance and the capacitive coefficients of the system S will be normalized to $4\pi\varepsilon r$, the potential of the body V will be given in volts, and the linear dimensions of the body will be dimensionless (normalized by the unit radius r).

Let the system S comprise a sphere S_1 of unit radius $R_1 = 1$ and a cylinder S_2 with a base radius $R_2 = 2$ and a height $h = 2$. A variety of the mutual arrangements of S_1 and S_2 can be divided into two cases: $S = S_1 \cup S_2$ is, or is not, a surface of revolution. In the former case, the center of the sphere S_1 lies on the axis of revolution of the cylinder S_2. Then, for $H = 2$ (H is the spacing between S_1 and S_2), the capacitance $\tilde{C}$ and the capaci-

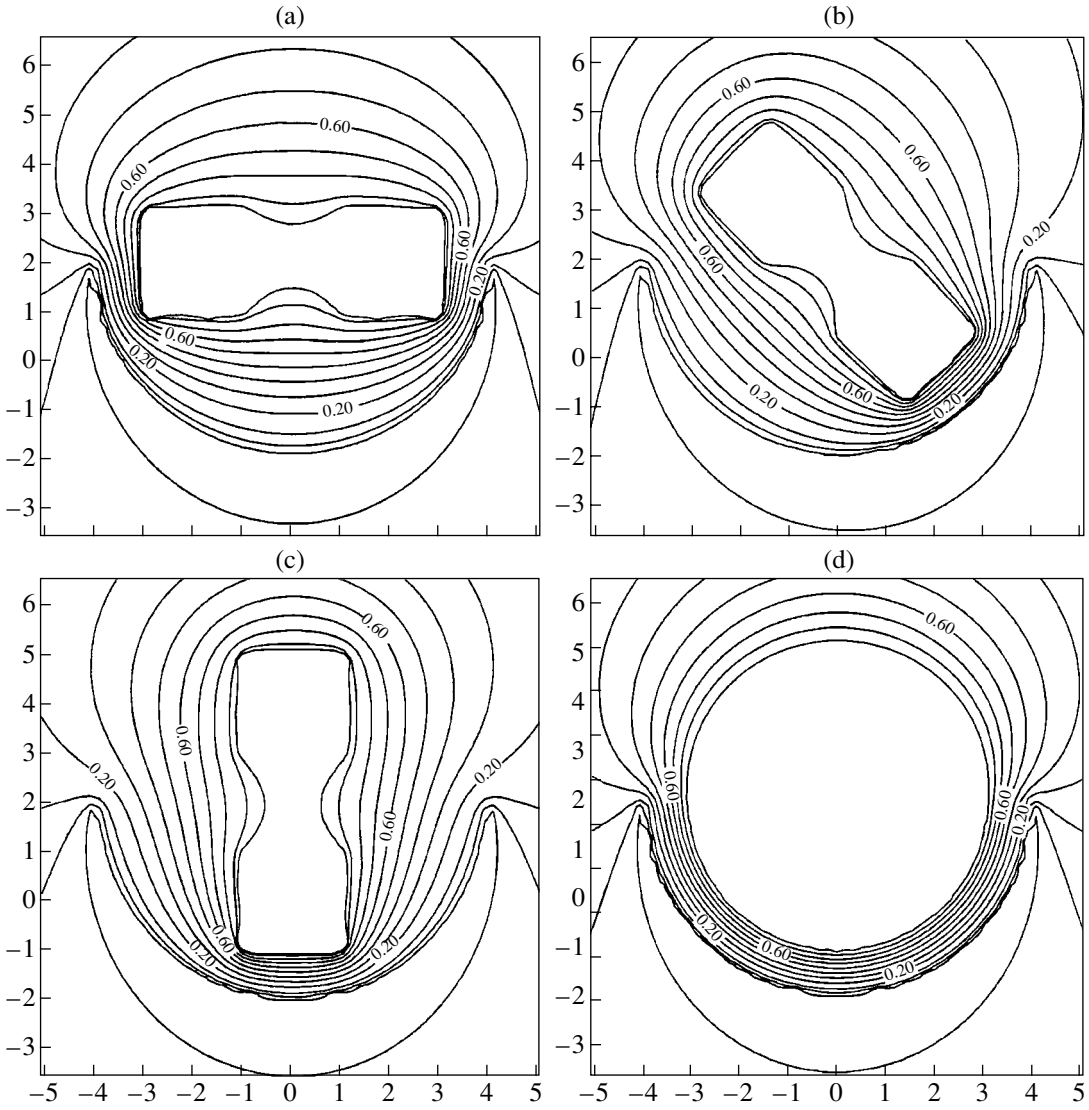


Fig. 3.

tive coefficients $c = \|c_{ij}\|$ ($i, j = 1, 2$) found by calculation are

$$\tilde{C} = 0.974219, \quad c = \begin{bmatrix} 1.14477 & -0.521235 \\ -0.521236 & 2.17735 \end{bmatrix}.$$

In the plane containing the axis of revolution of S (Fig. 4a), the electric potential for the system considered takes the maximal value (0.99) at $V_1 = V_2 = 1$ (V_1 and V_2 are the potentials of the screens from S_1 and S_2, respectively). Therefore, since the bodies S_1 and S_2 have equipotential volumes (according to the internal field criterion), the error of calculating the electrostatic field of S does not exceed 1%.

Let a point M lie on the axis of revolution of the system S and bisect the height of the cylinder S_2. We incline the cylinder in such a way that its axis crosses the axis of revolution of the system S at the point M at an angle of 45°. The resulting system S', $S' = S_1 \cup S_2$, is no longer a surface of revolution. For this system, the calculated capacitance $\tilde{C}$ and capacitive coefficients c_{ij} are as follows:

$$\tilde{C} = 1.01894, \quad c = \begin{bmatrix} 1.18618 & -0.571091 \\ -0.570945 & 2.21928 \end{bmatrix}.$$

Figure 4b shows the electric field potential of S' at $V_1 = V_2 = 1$ in the plane containing the axis of revolution of the cylinder S_2 and the center of the sphere S_1. As before, the maximal visualized value of the potential equals 0.99. Thus, for the nonaxisymmetric system S', the computational error remains within 1%.

Now, we pass to a system comprising a sphere S_1 of radius $R_1 = 1$ and a cone S_2 with a base radius $R_2 = 2$ and

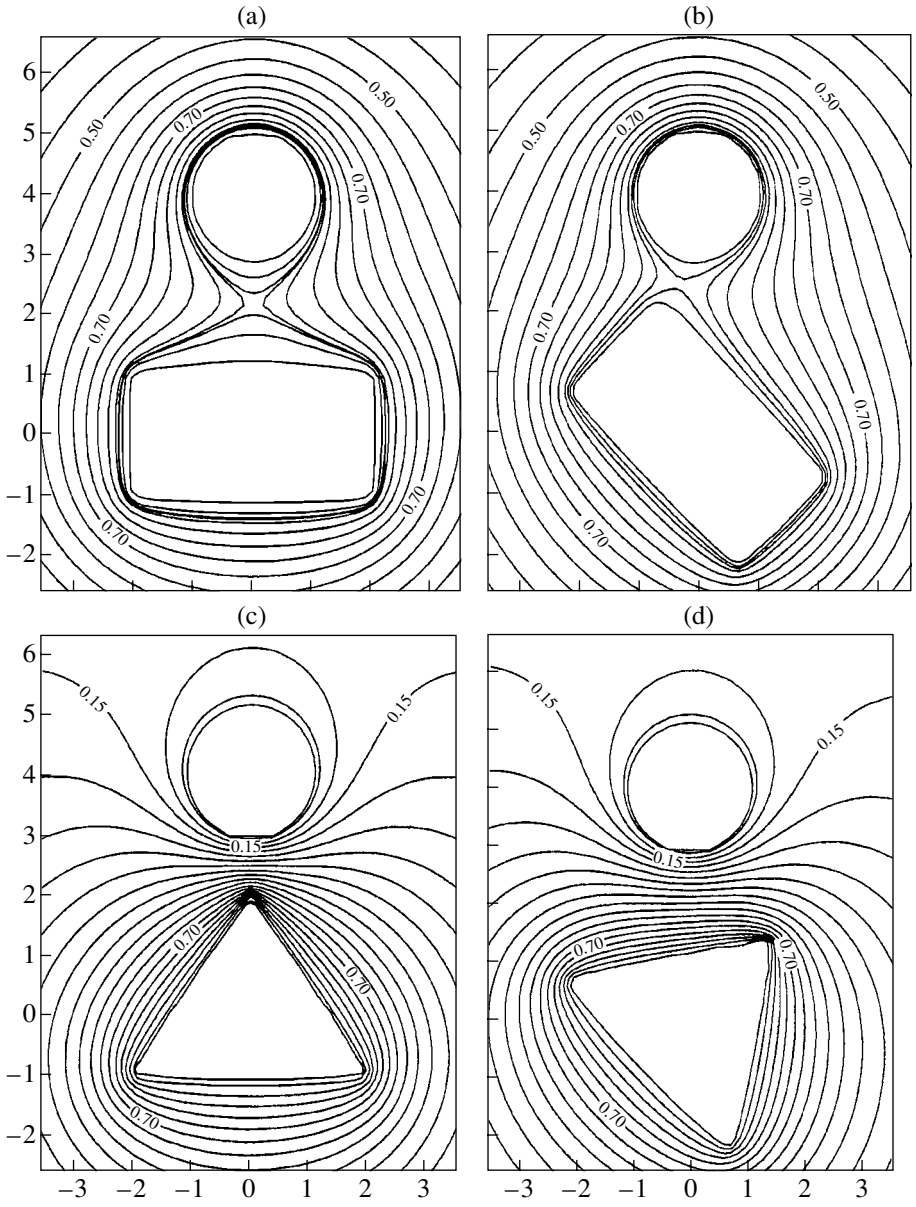


Fig. 4.

a height $h = 3$. It is assumed that $S = S_1 \cup S_2$ is a surface of revolution with a spacing $H = 1$ between S_1 and S_2. For this system, the capacitance $\tilde{C}$ and the capacitive coefficient matrix $c = \|c_{ij}\|$ $(i, j = 1, 2)$ are as follows:

$$\tilde{C} = 0.881883, \quad c = \begin{bmatrix} 1.12213 & -0.419433 \\ -0.419433 & 1.77205 \end{bmatrix}.$$

Equipotential lines of the entire system S in the plane containing its axis of revolution are depicted in Fig. 4c for $\varphi \in [0.01, 0.99]$ with $V_1 = 0$ and $V_2 = 1$ applied to the screens of S_1 and S_2, respectively. In this case, too, the error involved in the electric field computation is no more than 1%.

At a distance of $1/3h$ from the bottom of the cone, we fix a point M on its height and rotate the cone so that its axis crosses the axis of revolution of S at the point M at an angle of 45°; that is, we pass from the initial system to system S', $S' = S_1 \cap S_2$, which does not posses symmetry of rotation. For this system, the capacitance and the capacitive coefficient matrix are as follows:

$$\tilde{C} = 0.87989, \quad c = \begin{bmatrix} 1.11655 & -0.42029 \\ -0.420219 & 0.77272 \end{bmatrix}.$$

Figure 4d shows the cone-inclination-related distortion of the isolines of the electrostatic field for the system S (Fig. 4c) in the plane containing the axis of the

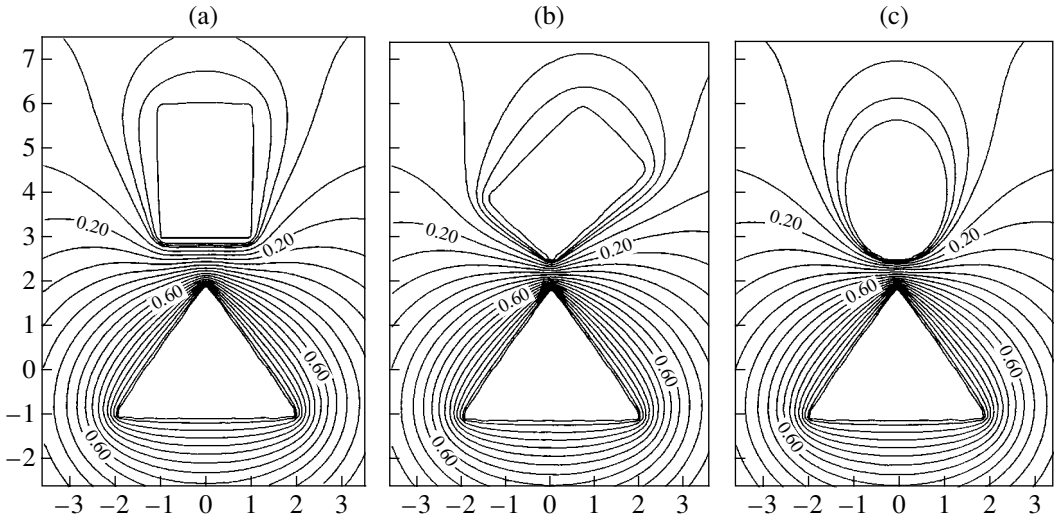


Fig. 5.

cone and the center of the sphere. In going from the system S to the system S', the computational error again does not exceed 1%.

Consider the third two-component system S consisting of the cone S_2 described above and a cylinder S_1 with a base radius $R_1 = 1$ and a height $h_1 = 3$ under the assumption that $S = S_1 \cup S_2$ is a surface of revolution. The spacing between S_1 and S_2 is set equal to $H = 1$. For this system, the capacitance and the capacitive coefficient matrix are as follows:

$$\tilde{C} = 1.12645, \quad c = \begin{bmatrix} 1.82498 & -0.557362 \\ -0.557362 & 1.59009 \end{bmatrix}.$$

Equipotential lines of the entire system S in the plane containing its axis of revolution are depicted in Fig. 5a for $\varphi \in [0.005, 0.995]$ with $V_1 = 0$ and $V_2 = 1$ applied to S_1 and S_2, respectively. In these calculations, the internal field criterion is fulfilled with an accuracy within 0.5%.

At a distance of $1/3h$ from the cylinder base nearest to the cone, we fix a point M on the cylinder axis and shift the cylinder S_1 so that the axes of the bodies intersect at the point M, making an angle of 45°. For such a system S', $S' = S_1 \cup S_2$, the capacitance and the capacitive coefficient matrix are as follows:

$$\tilde{C} = 1.13217, \quad c = \begin{bmatrix} 1.82871 & -0.564409 \\ -0.564409 & 1.59497 \end{bmatrix}.$$

The inclination of the cylinder S_1 distorts electrostatic potential isolines for the system S (Fig. 5a), as shown by Fig. 5b for the plane containing the axes of revolution of the cylinder S_1 and cone S_2 (Fig. 5b). As in the case of the axisymmetric system S, the error in calculating the potential of S' is within 0.5%.

To this point, the axisymmetric electrostatic fields of the systems S have been compared with the 3D fields of the systems S' under the assumption that the equipotential lines of the latter are visualized in the planes of mirror symmetry. However, a fuller description of 3D fields through their equipotential lines requires the lines to be constructed at least in three orthogonal planes (in practice, the number of sections for a given system S' may be as high as several tens). By way of example, consider the section of the above system S' passing through the axis of revolution of the cone S_2 perpendicularly to the axis of revolution of the cylinder S_1, that is, the section normal to that considered above (Fig. 5b). While the section of the cone remains an isosceles triangle with a base of 4 and a height of 3, the section of the cylinder S_1 becomes an ellipse instead of a 2×3 rectangle. Figure 5c shows isolines for the electrostatic field of the system S' in this plane for the same potential values as in Fig. 5b. It is easy to see that the error involved in field calculation using the internal field criterion does not exceed 0.5%, with the 0.005 isoline being an ellipse.

Finally, we will consider a surface of revolution $S = S_1 \cup S_2$, where S_2 is the surface generated by two cones with a mutual base of radius $R_2 = 0.5$ and heights of $h_2' = 1$ and $h_2'' = 0.5$ and S_1 is a cone with a base radius $R_1 = 2$ and a height $h_1 = 5$. The cone S_1 screens the surface S_2, which is located at a distance $H = 1.5$ from the cone base. The radial section of the system S is depicted in Fig. 2b. The capacitance and the capacitive coefficient matrix for this system are as follows:

$$\tilde{C} = 0.696219, \quad c = \begin{bmatrix} 2.58405 & -0.696219 \\ -0.696219 & 0.696219 \end{bmatrix}.$$

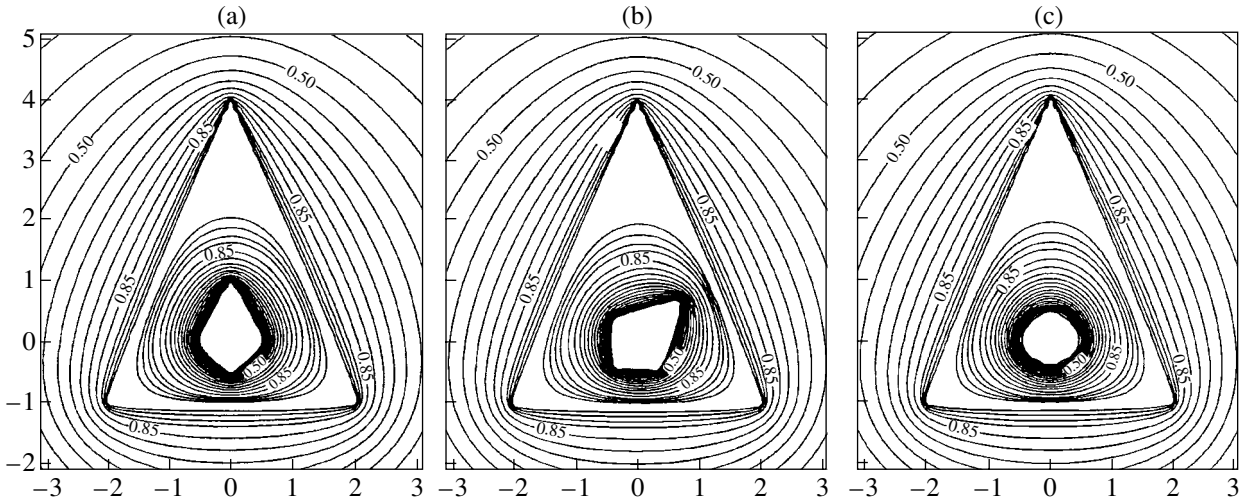


Fig. 6.

It is noteworthy here that the relationships

$$\tilde{C} = -c_{12} - c_{21} = c_{22} \tag{22}$$

are fulfilled with computer accuracy.

Equipotential lines for the field produced by the system S with potentials $V_1 = 1$ and $V_2 = 1$ applied to the screens S_1 and S_2 are shown in Fig. 6a in the radial section. The error is within 0.5%.

Now, we replace the axisymmetric system S by a system $S' = S_1 \cup S_2$ where the axis of revolution of the body S_2 crosses the height of the cone S_1 at an angle of 45°. The point of intersection is 1/5h distant from the cone base. The capacitance and the capacitive coefficient matrix for the system S' are as follows:

$$\tilde{C} = 0.707117, \quad c = \begin{bmatrix} 2.58854 & -0.707331 \\ -0.706919 & 0.0707117 \end{bmatrix}.$$

Here, equality (23) is fulfilled to four-place accuracy.

The variation of the electrostatic field structure (isolines in Fig. 6a) due to the transition from S to S' is shown in Figs. 6b and 6c. Namely, Fig. 6b depicts equipotential lines of the system S' in the plane containing the axes of revolution of the surfaces S_1 and S_2, while Fig. 6c demonstrates these lines in the plane passing through the axis of revolution of the cone S_1 orthogonal to the axis of revolution of the body S_2. According to the internal field criterion, the error of calculating the potential is no more than 0.5% for both systems (S' and S).

To conclude this section, we note that a computational accuracy of 1% was achieved when the surfaces considered above were approximated by ~10^4 boundary components.

(4) Consider a system $-S = S_1 \cup S_2 \cup S_3$ comprising three basic components. S_1 is a rod of length $h_1 = 4$ with a 2×2 cross section, S_2 is a torus with the generator in the form of an isosceles triangle with a base $\hat{h}_2 = 2$ and a height $h_2 = 1.7$ (the distance from the axis of revolution of the torus to the generator is $R_2 = 2.5$), and S_3 is a cylinder of height $h_3 = 1$ and a base radius $R_3 = 1$. The base of the cylinder has a hole of radius $R_{3d} = 0.5$. The axes of the surfaces of revolution S_2 and S_3 intersect at a right angle at the center of the square rod S_1. The spacing between S_1 and S_2 is $H_1 = 1.5$ and that between S_2 and S_3, $H_2 = 0.8$. The section P_1 passing through the axes of revolution of the surfaces S_2 and S_3 is depicted in Fig. 2c. For the rod S_1, this section is a 2×4 rectangle; for the surface S_2, two equal isosceles triangles with a base $\hat{h}_2 = 2$ and a height $h_2 = 1.7$ placed specularly symmetrically with respect to the rectangle at a distance $H_1 = 1.5$ from it; and for the cylinder S_3 with the hole, a 1×2 rectangle with a hole of size 1 on the side facing the rod. The section P_2 passing through the axis of revolution of the cylinder S_3 and the center of the rod S_1 (P_2 is orthogonal to the axis of revolution of the torus S_2) produces a 2×2 square and a ring with inner and outer radii $R_2' = 2.5$ and $R_2'' = 4.2$ that is centered at the center of the square. The section P_2 of the cylinder S_3 coincides with the section P_1.

The capacitive coefficient matrix $c = \|c_{ij}\|$ ($i, j = 1, 2, 3$) for such a system has the form

$$c = \begin{bmatrix} 3.28571 & -2.69617 & -0.0682739 \\ -2.68915 & 5.60708 & -0.669322 \\ -0.0672863 & -0.669852 & 1.21556 \end{bmatrix}.$$

Equipotential lines for the electrostatic field of the system S in the sections P_1 and P_2 are depicted in Figs. 7a and 7b, respectively, for potentials $V_1 = 1$, $V_2 =$

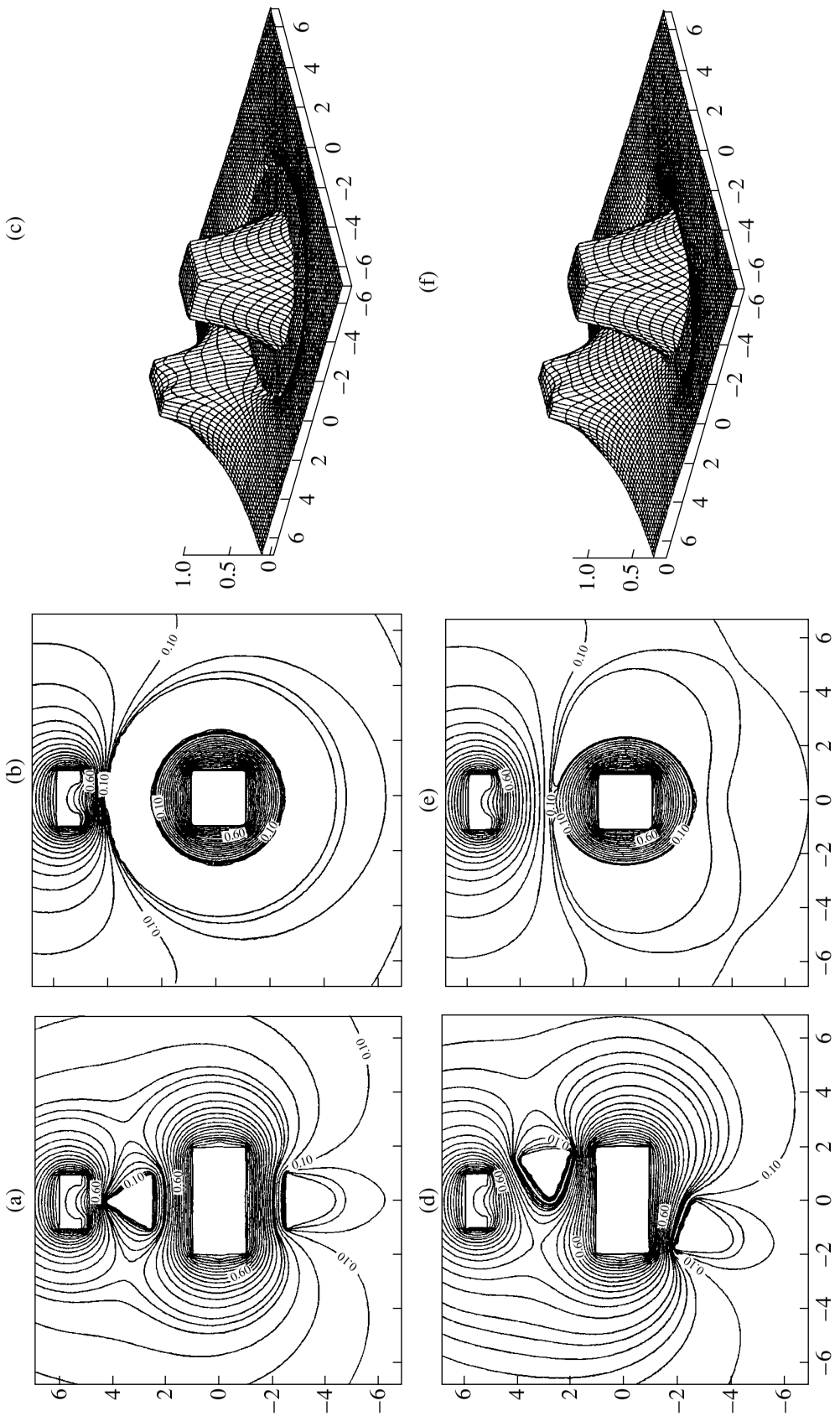


Fig. 7.

0, and $V_3 = 1$ applied to respective screens S_1, S_2, and S_3. Figure 7c shows the potential distribution in the section P_2. The error in potential calculation is within 1%, as demonstrated by Figs. 7a–7c.

Let us pass from system S to a system $S' = S_1 \cup S_2 \cup S_3$ where the axis of the cylinder with the hole and that of the torus intersect at the center of the rod, making an angle of 112°. Note that the planes P_1 and P_2 are the planes of mirror symmetry for the system S alone: the system S' has a single plane of mirror symmetry, P_1.

The matrix of the capacitive coefficients has the form

$$c = \begin{bmatrix} 3.41561 & -2.84179 & -0.0787125 \\ -2.85176 & 5.68682 & -0.563878 \\ -0.0777562 & -0.564894 & 1.13479 \end{bmatrix}.$$

Equipotential lines for the electrostatic field of the system S' in the sections P_1 and P_2 are depicted in Figs. 7d and 7e, respectively, for potentials $V_1 = 1$, $V_2 = 0$, and $V_3 = 1$ applied to respective screens S_1, S_2, and S_3. Figure 7f shows the potential distribution in the section P_2. In going from S to S', the error in potential calculation remains unchanged.

Note that a computational accuracy of 1% in calculating the field potentials for the systems S and S' was achieved when these multiply connected surfaces were approximated by $\approx 2 \times 10^4$ boundary components.

All the calculations were performed with the single software suite on a Pentium-Pro 200 MHz 128-MB RAM PC. The machine time per computation cycle at given screen potentials did not exceed 10 min.

REFERENCES

1. S. K. Demin, S. I. Safronov, and R. P. Tarasov, Zh. Tekh. Fiz. **68** (2), 97 (1998) [Tech. Phys. **43**, 222 (1998)].
2. S. K. Demin, S. I. Safronov, and R. P. Tarasov, Zh. Tekh. Fiz. **68** (7), 126 (1998) [Tech. Phys. **43**, 861 (1998)].
3. S. I. Safronov and R. P. Tarasov, Zh. Tekh. Fiz. **69** (6), 1 (1999) [Tech. Phys. **44**, 609 (1999)].
4. W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill, New York, 1950; Inostrannaya Literatura, Moscow, 1954).
5. S. K. Demin and R. P. Tarasov, Zh. Vychisl. Mat. Mat. Fiz. **29**, 1308 (1989).
6. E. V. Zakharov, S. I. Safronov, and R. P. Tarasov, Dokl. Akad. Nauk SSSR **314**, 589 (1990) [Sov. Phys. Dokl. **35**, 799 (1990)].
7. R. P. Tarasov, Zh. Vychisl. Mat. Mat. Fiz. **31**, 1515 (1992).
8. R. P. Tarasov, Zh. Vychisl. Mat. Mat. Fiz. **33**, 1815 (1993).
9. I. A. Zagorodnov and R. P. Tarasov, Zh. Vychisl. Mat. Mat. Fiz. **38**, 1303 (1998).
10. E. V. Zakharov, S. I. Safronov, and R. P. Tarasov, Zh. Vychisl. Mat. Mat. Fiz. **31**, 40 (1992).
11. R. P. Tarasov, Zh. Vychisl. Mat. Mat. Fiz. **39**, 943 (1999).
12. S. K. Demin and R. P. Tarasov, Mat. Model. **5** (7), 113 (1993).
13. E. V. Zakharov, S. I. Safronov, and R. P. Tarasov, Zh. Vychisl. Mat. Mat. Fiz. **33**, 1030 (1993).
14. E. V. Zakharov, S. I. Safronov, and R. P. Tarasov, Zh. Vychisl. Mat. Mat. Fiz. **38**, 734 (1998).
15. S. I. Safronov, Doctoral Dissertation (Moscow State Univ., Moscow, 2000).
16. G. Ch. Shushkevich, Zh. Tekh. Fiz. **68** (7), 1 (1998) [Tech. Phys. **43**, 743 (1998)].
17. E. V. Zakharov, S. I. Safronov, and R. P. Tarasov, Dokl. Akad. Nauk **367**, 457 (1999).
18. O. F. Antonenko, Vychisl. Sist., No. 12, 39 (1964).

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1083–1089. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 13–20.

ATOMS, SPECTRA, RADIATION

Complete Spin Polarization of 0.5- to 500-eV Electrons Elastically Scattered by Argon Atoms: Theoretical Consideration

V. I. Kelemen

Institute of Electronic Physics, National Academy of Sciences of Ukraine, Uzhgorod, 88000 Ukraine
e-mail: kelemen@iep.uzhgorod.ua

Received May 25, 2001; in final form, December 29, 2001

Abstract—Within the framework of a phenomenological real optical potential, the elastic scattering of 0.5- to 500-eV electrons by argon atoms is studied in view of spin–orbit interaction. The energy dependences of the angular positions of minima in differential cross sections are calculated and compared with those obtained in experiments. The energies and angular positions of four critical points where the cross sections reach their minimal values are determined. It is found that the deepest minimum of three high-angle ones (with angles >90°) is located at (38.25 eV, 140.67°). The energy and angular neighborhoods of these minima where the parameter of spin polarization, the so-called Sherman function, reaches its extreme values (+1 and –1) are determined.

INTRODUCTION

Over recent decades, elastic electron scattering by argon atoms has been thoroughly studied both experimentally and theoretically (see, e.g., reviews [1, 2]). Far fewer works have been devoted to electron polarization as a result of the scattering of an initially nonpolarized electron beam by argon atoms in the ground state.

The angular dependence of the spin polarization parameter $S(\Theta)$, the so-called Sherman function, was measured in [3–5] for energies of incident electrons from 10 to 150 eV in the angle range from 30° to 150°. In the theoretical works [6–8], the angular dependence of the parameter $S(\Theta)$ was calculated in the energy range of 3–300 eV. The maximum value of the spin polarization found in [3–8] did not exceed 20%.

At the same time, it is well known (see, e.g., [9]) that within small neighborhoods of the deepest minima of the differential cross sections (DCSs), the so-called critical minima, there are energies and angles at which the electron beam scattered can be completely, or almost completely, polarized. On the one hand, information about these energies and angles could help to measure experimentally the extreme values of the Sherman function (+1 and –1). On the other hand, both experimental techniques and theoretical models could be verified more efficiently at these points.

The aims of this paper are (i) to determine the critical minima of the DCSs for elastic electron scattering by argon atoms in the wide energy range of 0.5–500 eV and (ii) to calculate the energies and angles in the vicinity of these minima at which the polarization of scattered electrons could reach 100%.

THEORETICAL METHOD

Below, the variable phase method [10, 11] is used to calculate the cross sections and parameters of spin polarization. This method allows one to find the absolute values of the phase shifts $\delta_l^\pm \equiv \delta_l^{l \pm 1/2}$ of partial scattering amplitudes as the limiting values of the phase functions $\delta_l^\pm(r)$ at $r \longrightarrow \infty$. The phase functions are obtained from solving the phase equation

$$\frac{d\delta_l^\pm(r)}{dr} = -\frac{2V^\pm(r)}{k} \times [\cos\delta_l^\pm(r) j_l(kr) - \sin\delta_l^\pm(r) n_l(kr)]^2, \quad (1)$$

where j_l and n_l are the Riccati–Bessel functions (see, e.g., [10]), l is the orbital momentum, $k = (2E)^{1/2}$ is the momentum, E is the incident electron energy (hereafter, the atomic units $e = m_e = \hbar = 1$ are used), and $V^\pm(r)$ is the interaction potential of an incident electron with a target atom.

Using the phase shifts $\delta_l^\pm$, it is possible to find both the direct scattering amplitude,

$$f(\Theta) = \frac{1}{2ik}\sum_{l=0}^{\infty}\{(l+1)[\exp(2i\delta_l^+) - 1] + l[\exp(2i\delta_l^-) - 1]\}P_l(\cos\Theta), \quad (2)$$

and spin-flip scattering amplitude,

$$g(\Theta) = \frac{1}{2ik}\sum_{l=1}^{\infty}[\exp(2i\delta_l^-) - \exp(2i\delta_l^+)]P_l^1(\cos\Theta), \quad (3)$$

where Θ is the scattering angle; $P_l(\cos\Theta)$, Legendre polynomials; and $P_l^1(\cos\Theta)$, Legendre associated functions of first kind (see, e.g., [12]).

The elastic scattering DCS and the parameters of spin polarization of scattered electrons are expressed through the amplitudes f and g as [9]

$$\frac{d\sigma(\Theta)}{d\Omega} = |f(\Theta)|^2 + |g(\Theta)|^2,$$

$$S(\Theta) = i\frac{fg^* - f^*g}{|f|^2 + |g|^2}, \quad U(\Theta) = \frac{fg^* + f^*g}{|f|^2 + |g|^2}, \quad (4)$$

$$T(\Theta) = \frac{|f|^2 - |g|^2}{|f|^2 + |g|^2}.$$

In this paper, we are interested in energies and angles at which $S^2 = 1$; hence, since the polarization parameters are related to each other by the condition $S^2 + U^2 + T^2 = 1$, $U^2 = T^2 = 0$, and we will restrict our analysis to calculating the Sherman function $S(\Theta)$ and the DCS.

The phenomenological real optical potential

$$V^{\pm}(r, E) = V_s(r) + V_{s0}^{\pm}(r) + V_e(r, E) + V_p(r) \quad (5)$$

is used here as the interaction potential between an incident electron and a target atom.

The electrostatic potential $V_s(r)$ and the electron density in an argon atom (the latter is necessary to calculate the exchange potential $V_e(r, E)$) were found by analytical expressions in [13], where these parameters were determined with the Hartree–Fock method combined with the least-squares method.

Spin–orbit interaction is taken into account here via the potential [14]

$$V_{s0}^{+}(r) = \zeta^{\pm}(j, l)\frac{1}{r}\frac{dV_s}{dr}\frac{\alpha^2}{2 + \alpha^2[E - V_s(r)]}, \quad (6)$$

where $\zeta^+(j, l) = l/2$ at $j = l + 1/2$ and $\zeta^-(j, l) = -(l + 1)/2$ at $j = l - 1/2$, j is the total angular momentum of an electron, and α is the fine structure constant.

As in our previous papers [15, 16], we use the local exchange potential found in [17, 18] in the free electron gas approximation and the dipole polarization potential with an adjustable parameter (see either (2) in [15] or (10) in [16]) as the exchange potential $V_e(r, E)$ and the polarization potential $V_p(r)$, respectively. The ionization potential of an argon atom ($I = 15.76$ eV), which is used to calculate V_e, was taken from [19]; the dipole polarizability of an argon atom ($\alpha_d = 11.08a_0^3$), from [20]. Hereafter, a_0 is the Bohr radius of a hydrogen atom.

Since potential (5) has a singularity $r^2V^{\pm}(r) \longrightarrow r^2V_{s0}^{\pm}(r) \longrightarrow \text{const} = \zeta^{\pm}(j, l)$ for $r \longrightarrow 0$, the initial condition for the phase equation at the initial point of integration $r = \varepsilon$ has the form [11]

$$\delta_l^{\pm}(\varepsilon) = -q_l^{\pm}\frac{(k\varepsilon)^{2l+1}}{(2l+1)!!(2l-1)!!},$$
$$q_l^{\pm} = 1 + \frac{(2l+1)^2}{2\zeta^{\pm}} - \frac{(2l+1)^2}{2\zeta^{\pm}}\sqrt{1 + \frac{4\zeta^{\pm}}{(2l+1)^2}}. \quad (7)$$

The variable $q_l^{\pm}$ lies in the range $-1 < q_l^{\pm} < 1$. In the case of the attractive potential $V_{s0}^-(r)$ (i.e., when $\zeta^- = -(l + 1)/2$, ζ^- must satisfy the condition $\zeta^- > -(l + 1/2)^2$, resulting from the requirement that a particle not fall on the center. As is easily seen, this condition is satisfied for any $l \geq 1$.

To numerically solve the phase equation, we used, as in [15, 16], the fourth-order method with the Adams–Bashfort prediction and Adams–Moulton correction with modification [12]. The step size was controlled automatically. The number of partial waves involved in the calculations by the phase equation depended on the energy E. For example, six waves were taken into account at 1 eV and 26 waves, at 500 eV. At any energy, the phase shifts δ_l^+ and δ_l^- were calculated to $l = L$ at which $|\delta_l^+ - \delta_l^-| < 0.01\%$. For $l > L$, we assumed that $\delta_l^+ = \delta_l^-$. Finally, for partial waves with a phase shift smaller than 0.01 rad, the well-known formula for asymptotic phases was used to calculate δ_l. In the case when potential (5) decreases as $-\alpha_d/2r^4$ at large distances r, this formula has the form (see, e.g., [21])

$$\delta_l = \arctan\frac{\pi\alpha_d k^2}{(2l+3)(2l+1)(2l-1)}. \quad (8)$$

In order to determine the parameter in the polarization potential $V_p(r)$, we took advantage of the well-known fact that the Ramsauer–Townsend minimum is present near 0.3 eV in the low-energy electron scattering by argon atoms (see, e.g., [22, 23] and refs therein). For several values of this parameter, the energy dependence of the s-wave phase shift δ_0 was calculated at low energies. It was found that when the parameter was $1.99a_0$, the phase shift δ_0 first increased from 9.436 rad at $E = 0.001$ eV to 9.471 rad at $E = 0.07$ eV and then decreased to 9.426 rad $\approx 3\pi$ at $E = 0.3$ eV. Hence, for potential (5) with the adjustable parameter thus varying, there is a deep minimum in the s-wave partial cross section at 0.3 eV and accordingly the Ramsauer–Townsend minimum in the total cross section.

DISCUSSION

The DCSs of elastic electron scattering by argon atoms in the energy range E = 3–140 eV have two minima: low-angle (at angles <90°) and high-angle (at angles >90°) (Fig. 1). Since in our case the DCS values are defined by the direct scattering amplitude $f(\Theta)$ for almost all angles except the critical points, the angular positions of the minima in the DCS depend on the interference of partial wave terms in this amplitude. Note that the first three waves make a major contribution to the amplitude f in our energy range (for instance, at 30 eV and 40°, their contribution is greater than 90%, with the d-wave contribution being larger than 36%). Therefore, two minima (low-angle and high-angle) of $|f|^2$ are specified by the zeroes of the Legendre polynomial $P_2(\cos\Theta)$ at 54.7° and 125.3°. Comparison of our DCS with those calculated in [26] and experimentally measured in [2, 24, 25] (Fig. 1) shows that two theoretical results obtained by different approaches coincide with each other, unlike the experimental data, at any angle except the low-angle minimum.

Note that in [26] the partial phase shifts for l = 0–6 calculated at energies from 0 to 54.4 eV in the nonrelativistic approach are tabulated for 29 energies of an incident electron, whereas the DCS values are given for nine energies and angles ranging from 0° to 180° with a step of 5°. Therefore, in order to refine the angular positions of the minima and the DCS values in these minima, we calculated the cross section with a smaller angular step. For $l \geq 7$, the phase shifts were calculated by asymptotic formula (8) as in [26]. For intermediate energies, we approximated the phase shifts obtained in [26]. Thus, our energy dependences of both the angular positions and the minimum values of the DCSs from [26] are found with both tabulated and approximated phase shifts.

The energy dependences of the angular positions of the DCS low-angle and high-angle minima are shown in Figs. 2a and 2b. The experimental values of the angles of the minima for energies 3–100 eV are taken from both the tables and figures in [2] and from Table 1 in [27]; for energies lower than 3 eV and higher than 100 eV, from [28], where the data of 16 experiments are shown in the figures.

As follows from Figs. 2a and 2b, our results are in satisfactory agreement with most of the experimental data. The curve of low-angle minima (Fig. 2a) reaches a minimum value of 19.4° at 2.1 eV, passes through a maximum of 73.2° at 18.1 eV, and then gradually decreases to 59.4° at 133 eV. The calculation involving the phase shifts from [26] yields a minimum angle of 17° at 2.2 eV and a maximum value of 72.4° at 16.5 eV. For other energies, both calculations almost coincide. At energies higher than 133 eV, the low-angle minima flatten and DCS values at 90° do not exceed those for smaller angles.

In Fig. 2b, our calculation follows the trend of the experimental data: the curve of the high-angle DCS

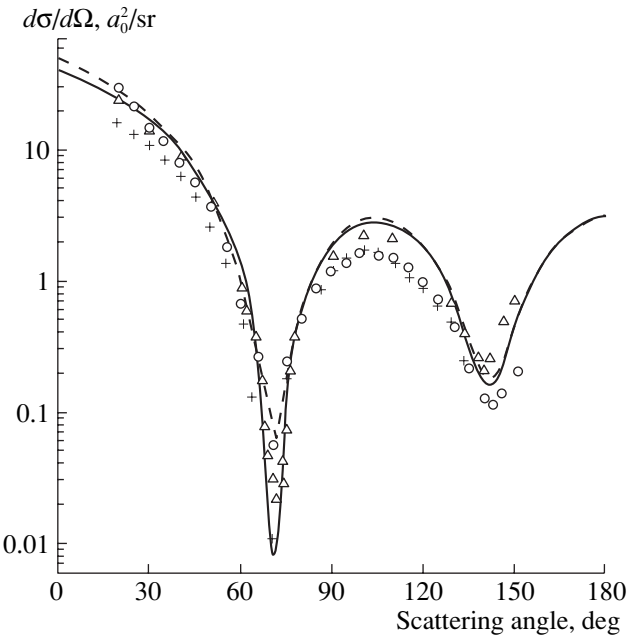


Fig. 1. Angular dependence of the differential cross section of elastic scattering of 30-eV electrons by atoms. Data points: ○ [2], + [24], and △ [25]. Solid line, calculation in [26]; dashed line, our results.

minima first gradually decreases to 119.7° at 9.6 eV, then reaches a maximum of 141.7° at 32.1 eV, and gradually decreases again to 96.6° at 500 eV. Starting from 3.5 eV, our calculation is also in good agreement with the theoretical results from [26].

Of these minima, critical ones can be conveniently found by plotting the energy dependence of the minimum DCS values. Figures 3a and 3b show such dependences calculated in this work. In Fig. 3a, the deepest low-angle minimum is located at (30.25 eV, 71.05°), where $|f|^2 = 6.5 \times 10^{-2}(a_0^2)$ and $|g|^2 = 7.4 \times 10^{-5}(a_0^2)$.

In Fig. 3b, the three lowest DCS values at the critical points H_1 = (8.44 eV, 119.89°), H_2 = (38.25 eV, 140.67°), and H_3 = (126.33 eV, 118.12°) are referred to as high-angle critical minima. At H_1, $|f|^2 = 5.9 \times 10^{-7}(a_0^2)$ and $|g|^2 = 1.7 \times 10^{-4}(a_0^2)$; at H_2, $|f|^2 = 1.8 \times 10^{-8}(a_0^2)$ and $|g|^2 = 1.05 \times 10^{-5}(a_0^2)$; and at H_3, $|f|^2 = 9.7 \times 10^{-9}(a_0^2)$ and $|g|^2 = 1.45 \times 10^{-5}(a_0^2)$. As can be seen, the deepest minimum is located at H_2.

When comparing the values of the critical points with experimental data, we had only two experimental works at our disposal (Table 1) where the critical minimum positions in the DCS of elastic scattering of intermediate-energy electrons by argon atoms are reported. In [29], three critical points were found for electrons of energies above 30 eV. In [2], the energy range from 30 to 60 eV and scattering angles from 60° to 150° were thoroughly studied to exactly locate critical points, and

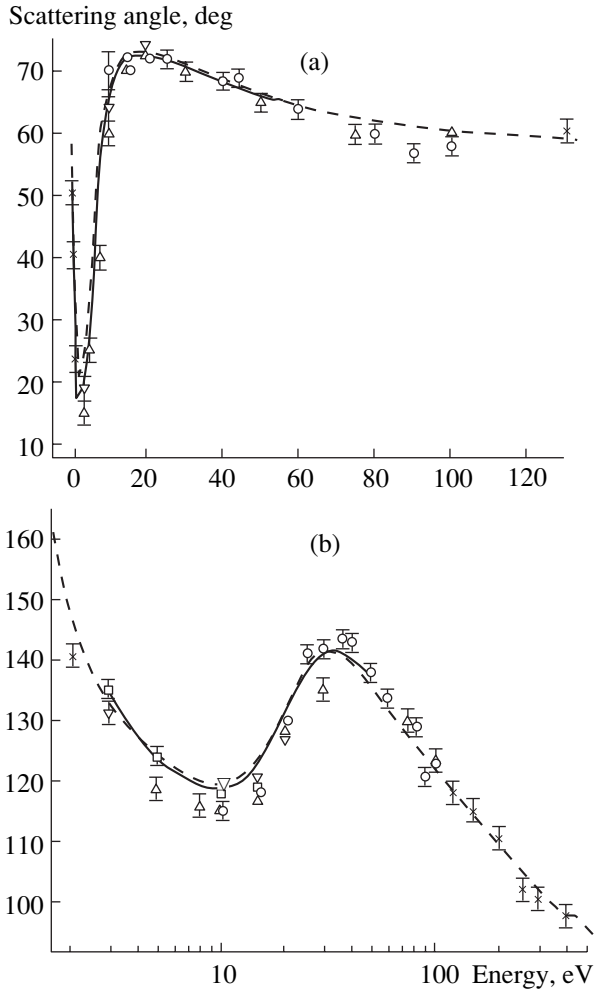


Fig. 2. Energy dependence of the angular positions for (a) low-angle and (b) high-angle minima of the differential cross sections of elastic electron scattering by atoms. Data points: ○ [2], △ [24], □ [27], × [28], and ▽ [29]. Solid line, calculation in [26]; dashed line, our results.

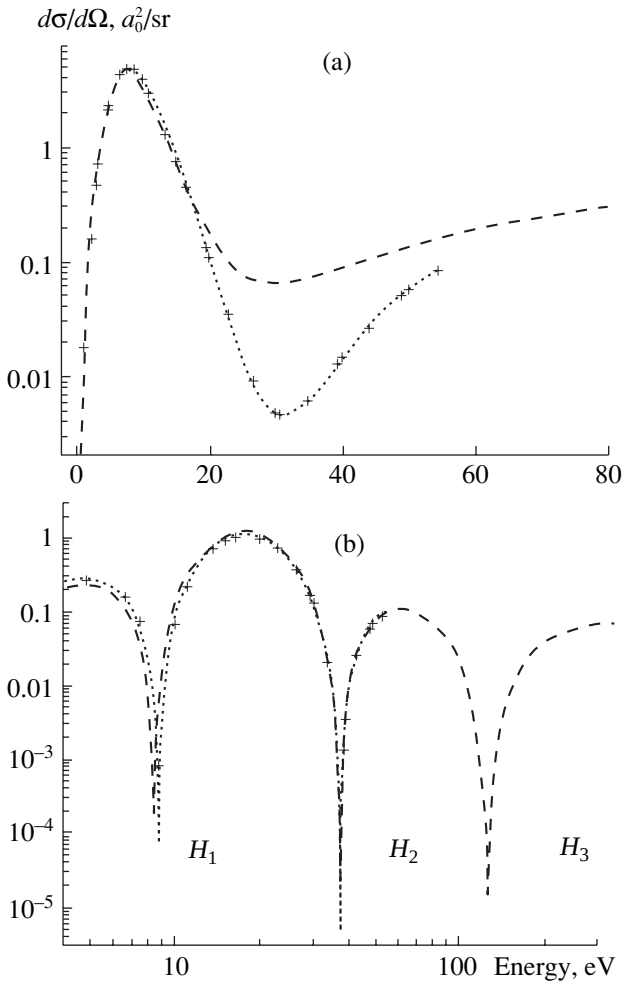


Fig. 3. Energy dependence of the (a) low-angle and (b) high-angle minima of the differential cross sections of elastic electron scattering by atoms. Our results are shown by the dashed line with the critical points H_1 = (8.44 eV, 119.89°), H_2 = (38.25 eV, 140.67°), and H_3 = (126.33 eV, 118.12°). The calculation uses (+) tabulated and (dots) approximated phase shifts from [26].

two were found. Comparison of the data in [2] and [29] shows that the positions of the low-angle critical minima differ in both angle (by 3°) and energy (by 15 eV). The angular positions of the high-angle critical minima coincide with each other, whereas the energies differ by 5 eV.

In [30], the angular dependences of the DCSs were calculated for energies and angles only in the vicinity of the two critical points found in [2]. The calculation was performed in the relativistic approximation with exact consideration of the exchange between an incident electron and electrons of a target atom and with regard for the *ab initio* polarization potential found by the relativistic method of polarized orbits. The positions of the critical minima found in [30] are in good agreement with the experiment [2]. However, there are differences in both the energy (by 2 eV) and angle (by 2°) for the high-angle minimum.

Table 1 shows that the theoretical results confirm the experimental data [2] for the angular position of the low-angle critical minimum and the energy position of the high-angle critical minimum. As for the differences in the energies of the low-angle minimum, Fig. 2a shows that the angular positions of the DCS low-angle critical minima for energies between 30 and 40 eV gradually vary from 71.1° to 68.5° and are within the experimental angular error.

From Table 1, it follows that in all the three theoretical works, the angular position of the high-angle minimum H_2 is found to be near 141°, whereas in the experimental works, the associated value is 143.5°. The energy of this minimum lies apparently between 37 and 39 eV. In any case, from Table 5 in [2] it follows that at

Table 1. Energy and angular positions of the low-angle (E_L, Θ_L) and high-angle (E_H, Θ_H) critical minima

E_L, eV	Θ_L, deg	E_H, eV	Θ_H, deg	Refs.
56.1 ± 0.7	65.73 ± 0.05	42.3 ± 0.9	143.8 ± 0.20	[29]
		132.3 ± 0.3	120.90 ± 0.05	
		126.1	118.9	
41.3 ± 0.02	68.5 ± 0.3	37.3 ± 0.02	143.5 ± 0.3	[2]
30.25	71.05	8.44	119.89	Our calculation
		38.25	140.67	
		126.33	118.12	
31	70.4	8.76	119.22	Phase shifts from [26]
		38.29	140.83	
39.3	68.5	39.5	141	[30]

143.5°, the experimental DCSs take minimum values between 37.3 and 39.3 eV.

As far as the critical point H_1 is concerned, note that the pronounced minimum in the DCS at 10 eV and near 120° is well known from available publications. For example, in our paper, the angular position of this minimum is equal to 119.68°; in the experimental works [2] and [24], 115°; in [27] and [31], 118° and 119°, respectively; and in the theoretical works [1] and [26], 115° and 120°, respectively. In view of the fact that the DCS values are usually reported with a step of 5°, one may infer that the minimum at 10 eV is located between 115° and 120°. Figure 3b shows that the low-energy dependence of the DCS minima found in this paper has a form similar to that constructed using the phase shifts from [26]. The difference in the positions of the critical minimum H_1 is only 0.3 eV and 0.7°.

Since the energy position of the critical point H_3 is greater than 100 eV, we can compare the position of this point only with experimental data from [29]. Note that the energy of H_3 coincides with the theoretical result (126.1 eV) in [29] and that the angular position of H_3 is less than the theoretical value in [29] by only 0.8° and less than the experimental one by 1.8°.

In order to find energies and angles at which the Sherman function $S(E, \Theta)$ may reach its extrema, we performed calculations in small neighborhoods of the three high-angle critical points, where $|g|^2 \geq |f|^2$. As was noted [9], the parameters of spin polarization are difficult to measure: the most thorough experiments determine the Sherman function with an error no smaller than ±5. Therefore, Table 2 lists energies and angles at which the function $S(E, \Theta)$ takes its extreme values: $S \approx 1$ for $S \in [0.90, 1]$ and $S \approx -1$ for $S \in [-1, -0.90]$.

From Table 2, the Sherman function reaches its maxima and minima in the neighborhoods of the critical points. We will refer to these neighborhoods as extremal neighborhoods. In the case of H_1, at 119.78°, $S \approx 1$ in the energy range [8.41, 8.44] eV; at 119.99°, $S \approx -1$ in the energy range [8.45, 8.48] eV. Hence, the sizes of this neighborhood are (0.07 eV, 0.21°). In a similar way, we find that the sizes of the extremal neighborhood for H_3 are (1.37 eV, 0.43°).

In contrast to the critical points considered above, the energy of the point H_2 (38.25 eV) does not belong to its extremal neighborhood. Figure 4 shows that $|S(38.25 \text{ eV}, \Theta| \leq 0.5$ (curve *4*) for any angle from the neighborhood of H_2. The Sherman function reaches its extreme values at other energies from this neighborhood; namely, $S \approx 1$ in the energy range of 38.12–38.20 eV within the angles 140.71°–140.74° (curves *1–3*) and $S \approx -1$ in the energy range of 38.29–38.17 eV within the angles 140.61°–140.63° (curves *5–7*). Consequently, it may so happen that it is insufficient to investigate only angular neighborhoods of the critical minima; it may also be necessary to determine their extremal energy neighborhoods in order to find the energies and angles of a complete spin polarization.

Since predicting the sizes of the extremal neighborhoods is a subject of theoretical study, the question arises as to how these sizes depend on the potential selected. In order to answer this question, let us express

Table 2. Extreme values of the function $S(E, \Theta)$ in the vicinity of the critical points H_1 = (8.44 eV, 119.89°), H_2 = (38.25 eV, 140.67°), and H_3 = (126.33 eV, 118.12°)

	E, eV	Θ, deg	S	E, eV	Θ, deg	S
H_1	8.41	119.78	0.93	8.45	119.99	−1.0
	8.43	119.78	1.0	8.46	119.99	−0.99
	8.44	119.78	0.99	8.48	119.99	−0.91
H_2	38.12	140.74	0.90	38.29	140.63	−0.90
	38.17	140.72	1.0	38.32	140.63	−1.0
	38.20	140.71	0.93	38.37	140.61	−0.92
H_3	125.63	118.34	−0.90	126.32	118.01	0.96
	126.07	118.26	−1.0	126.58	117.98	1.0
	126.31	118.23	−0.96	127.0	117.91	0.90

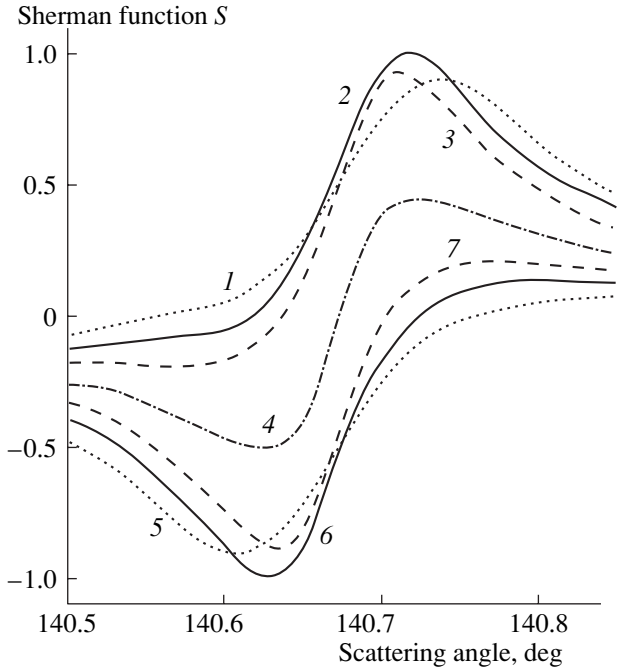


Fig. 4. Angular dependences of the Sherman function $S(E, \Theta)$ in the vicinity of the critical point H_2 for energies E = (*1*) 38.12, (*2*) 38.17, (*3*) 38.20, (*4*) 38.25, (*5*) 38.37, (*6*) 38.32, and (*7*) 38.29 eV.

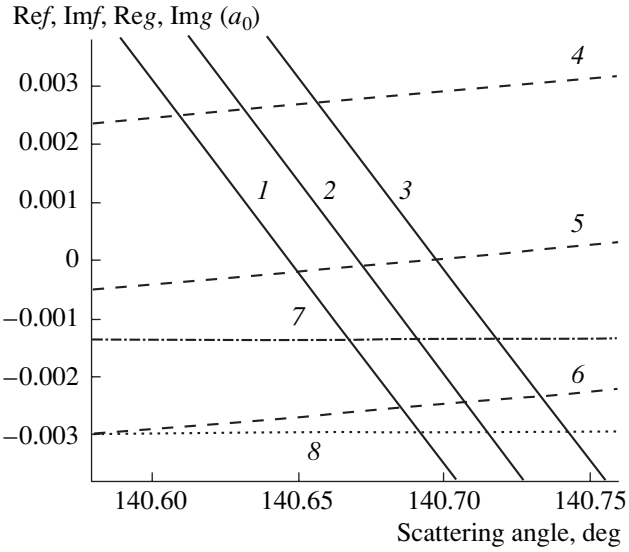


Fig. 5. Angular dependences of the real and imaginary parts of the amplitudes f and g in the vicinity of the critical point H_2. The energies E for Ref are (*1*) 38.32, (*2*) 38.25, and (*3*) 38.17; for Imf, (*4*) 38.17, (*5*) 38.25, and (*6*) 38.32; for Img, (*7*) 38.25; and for Reg, (*8*) 38.25 eV.

the Sherman function through the real and imaginary parts of the amplitudes f and g:

$$S = \frac{2(\mathrm{Re}\, f\, \mathrm{Im} g - \mathrm{Im}\, f\, \mathrm{Re} g)}{(\mathrm{Re}\, f)^2 + (\mathrm{Im}\, f)^2 + (\mathrm{Re} g)^2 + (\mathrm{Im} g)^2}. \qquad (9)$$

As follows from (9), $S = 1$ when Ref = Img and Imf = –Reg, and $S = -1$ when Ref = –Img and Imf = Reg. These two cases are demonstrated in Fig. 5 with the critical point H_2. In the extremal neighborhood of this point (see Table 2), $S = 1$ at 38.17 eV and 140.72° and $S = -1$ at 38.32 eV and 140.63°.

For all energies from this neighborhood, the values of the amplitude g coincide with those for 38.25 eV. Hence, we will assume that Re$g = -2.95 \times 10^{-3}(a_0)$ and Im$g = -1.36 \times 10^{-3}(a_0)$ throughout the neighborhood. For the energy 38.17 eV and the angle 140.72°, Ref = $-1.47 \times 10^{-3}(a_0)$ and Im$f = 2.97 \times 10^{-3}(a_0)$; i.e., indeed, Re$f \approx$ Img and Im$f \approx$ –Reg. For the energy 38.32 eV and the angle 140.63°, Ref = $1.17 \times 10^{-3}(a_0)$ and Imf = $-2.78 \times 10^{-3}(a_0)$; i.e., indeed, Re$f \approx$ –Img and Im$f \approx$ Reg. At intermediate energies from this neighborhood, e.g., at 38.25 eV, $|\mathrm{Re}\, g| > |\mathrm{Re}\, f|$ and $|\mathrm{Im}\, g| > |\mathrm{Im}\, f|$ (hence, $|g|^2 > |f|^2$) for any angle from the neighborhood. Thus, the sizes of the extremal neighborhood depend on both the value of $|g|^2$ and the width of the minimum of $|f|^2$ at the boundary of the neighborhood, where $|f|^2 \approx |g|^2 = 1.1 \times 10^{-5}(a_0^2)$.

From (3), it follows that the amplitude g depends on the spin–orbit splitting of the phase shifts $\Delta_l = \delta_l^- - \delta_l^+$. We will consider only the correction due to spin–orbit interaction of all relativistic corrections (the so-called semi-relativistic approach) and compare our results with the completely relativistic calculation, e.g., for an energy and an angle close to the critical point H_2, namely, for 40 eV and 141°. Using the phases from [6], we find for l = 1, 2, and 3 that $\Delta_1 = 0.0110$ rad, $\Delta_2 = 0.0017$ rad, $\Delta_3 = -2 \times 10^{-4}$ rad, and $|g|^2 = 8.5 \times 10^{-6}(a_0^2)$. Our results are $\Delta_1 = 0.0119$ rad, $\Delta_2 = 0.0016$ rad, $\Delta_3 = 7 \times 10^{-5}$ rad, and $|g|^2 = 1 \times 10^{-5}(a_0^2)$. Thus, the semi-relativistic approach yields almost the same values of Δ_l and $|g|^2$ as the completely relativistic one. Hence, the potentials used in this work satisfactorily describe the interaction of an incident electron with an argon atom at small distances from the nucleus and allow one to find reliable values of the amplitude g.

Figure 5 shows that in the neighborhood of the minimum H_2, Imf has a smaller slope than Ref and does not even cross the zero axis for several energies. Thus, the position of the minimum of $|f|^2$ is defined primarily by the point where Ref crosses the zero axis, while the minimum of $|f|^2$ depends chiefly on Imf.

The use of another optical potential, especially another polarization potential, will evidently result in another energy dependence of the phase shifts and, hence, another position of the DCS critical minimum. At the same time, Ref and Imf calculated with the approximated phase shifts from [26] behave in the same way as those in Fig. 5 within a small neighborhood of the critical minimum (38.29 eV, 140.83°). Namely, Ref

steeply crosses the zero axis. The crossing angle and, accordingly, the angular position of the DCS minimum shift toward smaller angles (see also Fig. 2b) with increasing energy. Having a substantially smaller slope than Ref, Imf becomes negative as the energy grows. The width of the DCS minimum, where $|f|^2 \approx 1.1 \times 10^{-5}(a_0^2)$, is 0.1° for 38.29 eV, i.e., equals our value for 38.25 eV.

For the high-angle critical minimum H_1, the calculations with approximated phase shifts from [26] yield the same behavior of Ref and Imf as that found in this paper. The width of the DCS minimum, where $|f|^2 \approx |g|^2 = 1.7 \times 10^{-4}(a_0^2)$, is 0.05° for 8.76 eV, coinciding with our result for 8.44 eV.

Since in a small neighborhood of the low-angle critical minimum, both Imf and Ref steeply cross the zero axis, the minimum values of $|f|^2$ equally depend on Imf and Ref. Figures 1 and 2a show that our angular positions of the DCS low-angle minima at energies from 20 to 50 eV perfectly agree with those obtained in [26]. However, as follows from Fig. 3a, at these energies, the DCS low-angle minima at the crossing points of Ref and Imf turn out to be more sensitive to small differences in the energy dependences of the phase shifts calculated in [26] and in this paper than the high-angle minima (Fig. 3b). Therefore, we propose to use the extremal neighborhoods of high-angle critical points for experimentally observing complete spin polarization of electrons scattered by argon atoms.

In conclusion, note that the positions of our DCS minima are in good agreement with the experimental data. The energy and angular positions of the minimum H_2 perfectly agree with the experiment [2]; those of the minima H_1 and H_3, with other theoretical calculations. Our energies and angles of complete spin polarization of scattered electrons in the neighborhood of high-angle minima may be employed in further polarization experiments. Even if the positions of DCS critical minima are refined in subsequent experiments, the extremal neighborhoods of critical points found in this article may be used for estimating the energies and angles of complete spin polarization of scattered electrons provided that the condition $|g|^2 \geq |f|^2$ is satisfied in the neighborhoods of the minima.

ACKNOWLEDGMENTS

The author is grateful to B. Marinković (Belgrade) for showing interest in this work and E. Remeta for fruitful discussions.

REFERENCES

1. S. N. Nahar and J. M. Wadehra, Phys. Rev. A **35**, 2051 (1987).
2. R. Panajotović, D. Filipović, B. Marinković, *et al.*, J. Phys. B **30**, 5877 (1997).
3. J. Mehr, Z. Phys. **198**, 345 (1967).
4. K. Schackert, Z. Phys. **213**, 316 (1968).
5. M. J. M. Beerlage, Zhou Qing, and M. J. van der Wiel, J. Phys. B **14**, 4627 (1981).
6. J. E. Sienkiewicz and W. E. Baylis, J. Phys. B **21**, 885 (1988).
7. S. N. Nahar and J. M. Wadehra, Phys. Rev. A **43**, 1275 (1991).
8. S. P. Khare and D. Raj, J. Phys. B **26**, 4807 (1993).
9. J. Kessler, Adv. At., Mol., Opt. Phys. **27**, 81 (1991).
10. F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York, 1967; Mir, Moscow, 1972).
11. V. V. Babikov, *Variable Phase Method in Quantum Mechanics* (Nauka, Moscow, 1988).
12. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1961; Nauka, Moscow, 1977).
13. T. G. Strand and R. A. Bonham, J. Chem. Phys. **40**, 1686 (1964).
14. R. Cowan, *The Theory of Atomic Structure and Spectra* (University of California Press, Berkeley, 1981).
15. V. I. Kelemen, E. Yu. Remeta, and E. P. Sabad, Zh. Tekh. Fiz. **61** (2), 46 (1991) [Sov. Phys. Tech. Phys. **36**, 150 (1991)].
16. V. I. Kelemen, E. Yu. Remeta, and E. P. Sabad, J. Phys. B **28**, 1527 (1995).
17. S. Hara, J. Phys. Soc. Jpn. **22**, 710 (1967).
18. S. Sur and A. S. Ghosh, Indian J. Phys. B **57**, 67 (1983).
19. A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions* (Énergoatomizdat, Moscow, 1986; Springer-Verlag, Berlin, 1985).
20. T. M. Miller and B. Bederson, in *Advances in Atomic and Molecular Physics*, Ed. by D. R. Bates and B. Bederson (Academic, New York, 1977), Vol. 13, pp. 1–55.
21. P. G. Burke, *Potential Scattering in Atomic Physics* (Plenum, New York, 1977; Atomizdat, Moscow, 1980).
22. L. S. Frost and A. V. Phelps, Phys. Rev. **136**, A1538 (1964).
23. Yu. K. Gus'kov, R. V. Savvov, and V. A. Slobodyanyuk, Zh. Tekh. Fiz. **48**, 277 (1978) [Sov. Phys. Tech. Phys. **23**, 167 (1978)].
24. S. K. Srivastava, H. Tanaka, A. Chutjian, *et al.*, Phys. Rev. A **23**, 2156 (1981).
25. J. F. Williams and B. A. Willis, J. Phys. B **8**, 1670 (1975).
26. R. P. McEachran and A. D. Stauffer, J. Phys. B **16**, 4023 (1983).
27. J. E. Furst, D. E. Golden, M. Mahgerefteh, *et al.*, Phys. Rev. A **40**, 5592 (1989).
28. D. M. Filipović and B. Marinković, in *Proceedings of the XX International Conference on Physics of Electronic and Atomic Collisions, Vienna, 1997*, p. TH 005.
29. J. Kessler, J. Liedtke, and C. B. Lukas, in *Proceedings of the International Summer School and Symposium on the Physics of Ionized Gases, Dubrovkik, 1976*, Ed. by B. Navinšek (J. Stefan Institute, Ljubljana, 1976), p. 61.
30. J. E. Sienkiewicz, S. Telega, and V. Konopińska, in *Proceedings of the 7th European Conference on Atomic and Molecular Physics (ECAMP), Berlin, 2001*, p. 68.
31. J. F. Williams, J. Phys. B **12**, 265 (1979).

Translated by M. Fofanov

Technical Physics, Vol. 47, No. 9, 2002, pp. 1090–1097. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 21–28.

GASES AND LIQUIDS

Instability of a Charged Spherical Drop Moving Parallel to an External Electrostatic Field

V. A. Koromyslov and A. I. Grigor'ev

Yaroslavl State University, Sovetskaya ul. 14, Yaroslavl, 150000 Russia

e-mail: grig@uniyar.ac.ru

Received February 13, 2002

Abstract—It is shown that the pressure of electrostatic fields induced by the self-charge of a drop and by the polarization charge and aerodynamic pressure of a laminar gas flow around a moving charged drop acting simultaneously reduce the critical instability conditions for the surface of the drop. For these conditions, the spectrum of capillary oscillations of the drop is calculated. It is found that, at various values of the charge, field strength, and velocity of the drop, the vibrational instability of the drop surface develops through the interaction of different oscillation modes, namely, second and third, second and fourth, and third and fifth. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Various problems of technical physics, geophysics, and technology deal with charged drops moving in an external electrostatic field [1–3], which motivates analysis of their stability. According to experimental data [4], an air flow around a water-covered hailstone decreases the critical strength of an external electrostatic field capable of initiating a corona discharge. Theoretical calculations [5, 6] show that the motion of a charged drop in the environment makes it less stable against its self-charge because of a tangential discontinuity in the velocity field at the interface. A more general problem of capillary oscillations and stability of a charged drop moving in a uniform electrostatic field has not yet been solved, although it is of great interest.

(1) We will find critical instability conditions for the surface of a charged spherical drop of a nonviscous perfectly conducting incompressible liquid of density ρ_2. The drop, having a radius R and a charge Q, moves in a uniform electrostatic field of strength $\mathbf{E}_0$ with a constant velocity $\mathbf{U}$ through a perfect incompressible dielectric fluid of density ρ_1 and permittivity ε. The interfacial tension is σ, and the direction of the drop is assumed to be aligned with the field.

Wave motions in the environment and in the drop are assumed to be harmonic and potential with the velocity potentials Ψ_1 and Ψ_2, respectively. A perturbation $\xi(\mathbf{r}, t)$ of the interface caused by the thermal motion of molecules is assumed to be axisymmetric and small, that is, $|\xi(\Theta, t)|/R \ll 1$. The analysis is performed in the spherical coordinate system with the origin at the center of the drop in the linear approximation with respect to $|\xi/R|$.

The assumption that the drop is spherical greatly facilitating further calculations, is based on the following reasoning. It is known [7] that a stagnant drop of a conductive liquid placed in a uniform electrostatic field takes the shape of a spheroid extended along the field. The pressure of a laminar flow around the drop acts in the opposite direction, that is, flattens the drop, because a reduction of the pressure due to an increase in the velocity is the greatest near the equatorial area of the drop according to the Bernoulli equation. Thus, when the drop moves collinearly with the field strength, these two tendencies may balance and the drop will be near-spherical.

To simplify calculations, let us introduce dimensionless variables such that $R = 1$, $\sigma = 1$, and $\rho_2 = 1$. Then, the rest of the variables involved in the problem (designated as usual) will be expressed in units of their characteristic values:

$$r_* = R, \quad t_* = R^{3/2}\rho_2^{1/2}\sigma^{-1/2},$$

$$U_* = R^{-1/2}\rho_2^{-1/2}\sigma^{1/2},$$

$$P_* = R^{-1}\sigma, \quad Q_* = R^{3/2}\sigma^{1/2},$$

$$E_* = R^{-1/2}\sigma^{1/2}, \quad \rho_1/\rho_2 \equiv \rho.$$

The perturbed surface of the drop is described by the equation

$$r(\Theta, t) = 1 + \xi(\Theta, t).$$

The mathematical statement of the problem has the form

$$\Delta\Phi = 0, \quad \Delta\Psi_i = 0 \ (i = 1, 2), \tag{1}$$

$$r = 0: \Phi = \text{const}, \quad \nabla\Psi_2 = 0, \tag{2}$$

$$r = 1 + \xi: \frac{\partial\xi}{\partial t} + \frac{1}{r^2}\frac{\partial\Psi_1}{\partial\Theta}\frac{\partial\xi}{\partial\Theta} = \frac{\partial\Psi_1}{\partial r}, \tag{3}$$

1063-7842/02/4709-1090$22.00 © 2002 MAIK "Nauka/Interperiodica"

$$\frac{\partial \xi}{\partial t} = \frac{\partial \Psi_2}{\partial r}, \quad (4)$$

$$-\rho\frac{\partial \Psi_1}{\partial t} + \frac{\partial \Psi_2}{\partial t} - \frac{1}{2}\rho(\nabla\Psi_1)^2 - P_E + P_\sigma = 0, \quad (5)$$

$$r \longrightarrow \infty: \Phi \longrightarrow -\mathbf{E}_0 \cdot \mathbf{r}, \quad \nabla\Psi_1 \longrightarrow \mathbf{U}. \quad (6)$$

Here, $\Phi(\mathbf{r}, t)$ is the electric field potential; $\Psi_1(\mathbf{r}, t)$ and $\Psi_2(\mathbf{r}, t)$ are the velocity field potentials in the environment and in the drop, respectively; $\mathbf{n}_z \parallel \mathbf{E}_0$; $P_\sigma = P_\sigma^0 + P_\sigma(\xi)$ is the pressure exerted by the surface tension forces; $P_E = P_E^0 + P_E(\xi)$ is the pressure of the electric field on the surface of the drop (see Appendix); and Δ is Laplacian. The quantities P_E^0 and P_σ^0 are of the zeroth order of smallness in ξ/R, while $P_E(\xi)$ and $P_\sigma(\xi)$ are of the first order of smallness in $|\xi/R|$.

(2) We will seek a solution to Laplace equations (1) for the velocity potentials $\Psi_1(\mathbf{r}, t)$ and $\Psi_2(\mathbf{r}, t)$ subject to boundary conditions (2) and (6). The perturbation $\xi(\Theta, t)$ will be sought in the form of the expansion in normalized axisymmetric spherical functions $Y_n(\mu)$:

$$\Psi_1(\mathbf{r}, t) = \varphi(\mathbf{r}) + \sum_{n=0}^{\infty} A_n r^{-(n+1)} Y_n(\mu)\exp(St),$$

$$\Psi_2(\mathbf{r}, t) = \sum_{n=0}^{\infty} B_n r^n Y_n(\mu)\exp(St), \quad (7)$$

$$\xi(\mathbf{r}, t) = \sum_{n=2}^{\infty} Z_n Y_n(\mu)\exp(St),$$

where $\mu \equiv \cos\Theta$; S is the complex frequency; the expansion amplitudes A_n, B_n, and Z_n are constant coefficients of the same order of smallness; and $\varphi(\mathbf{r})$ is the velocity field potential in the environment near the unperturbed spherical surface.

For ξ, the summation in the expansion starts with $n = 2$, because the zeroth mode corresponds to radial vibrations that are impossible in an incompressible liquid; consequently, $Z_0 = 0$. The first mode corresponds to the translational motion of the drop; in the frame of reference related to the center of mass of the drop, $Z_1 = 0$.

The gradient of $\varphi(\mathbf{r})$ defines the velocity field in an undisturbed flow of a nonviscous fluid near a sphere [8]:

$$\mathbf{V} = \nabla\varphi = -\frac{R^3}{2r^3}[3\mathbf{n}(\mathbf{U}\cdot\mathbf{n}) - \mathbf{U}] + \mathbf{U},$$

where $\mathbf{n}$ is the unit vector normal to the spherical surface of the drop.

According to the aforesaid, the direction of the velocity $\mathbf{U}$ coincides with the unit vector $\mathbf{n}_z$ in the Cartesian coordinate system. Taking into account that $\mathbf{e}_z \equiv \mathbf{e}_r\cos\Theta - \mathbf{e}_\Theta\sin\Theta$ and $\mathbf{n} \equiv \mathbf{e}_r$, we write an expression for the liquid flow velocity near a spherical drop in the form

$$\mathbf{V} = U\cos\Theta\left[1 - \frac{R^3}{r^3}\right]\mathbf{e}_r - U\sin\Theta\left[1 + \frac{R^3}{2r^3}\right]\mathbf{e}_\Theta. \quad (8)$$

Substituting expressions (7) and (8) into boundary conditions (3)–(5) and using the recurrent relation [9]

$$\sin\Theta\frac{\partial Y_n}{\partial\Theta} = \frac{n(n+1)}{\sqrt{(2n+1)(2n+3)}}Y_{n+1} - \frac{n(n+1)}{\sqrt{(2n+1)(2n-1)}}Y_{n-1},$$

we obtain for $r = 1$

$$\sum_n SZ_nY_n(\mu)\exp(St) - \frac{3}{2}U\sum_n Z_n\left(\frac{n(n+1)}{\sqrt{(2n+1)(2n+3)}}Y_{n+1} - \frac{n(n+1)}{\sqrt{(2n+1)(2n-1)}}Y_{n-1}\right)\exp(St) + \sum_n (n+1)A_nY_n(\mu)\exp(St) = 0, \quad (9)$$

$$\sum_n SZ_nY_n(\mu)\exp(St) = \sum_n nB_nr^nY_n(\mu)\exp(St), \quad (10)$$

$$-\rho S\sum_n A_nY_n\exp(St) + S\sum_n B_nY_n\exp(St) - P_E(\xi) + P_\sigma(\xi) + 3\rho U\sum_n A_n\left[\frac{n(n+1)}{\sqrt{(2n+1)(2n+3)}}Y_{n+1} - \frac{n(n+1)}{\sqrt{(2n+1)(2n-1)}}Y_{n-1}\right]\exp(St) = 0. \quad (11)$$

Here, it has been taken into account that, for $r = 1$,

$$\nabla\Psi_1 = \left[-\sum_n A_n(n+1)Y_n(\mu)\exp(St)\right]\mathbf{e}_r + \left[-\frac{3}{2}U\sin\Theta + \sum_n A_n\frac{\partial Y_n}{\partial\Theta}\exp(St)\right]\mathbf{e}_\Theta,$$

$$(\nabla\Psi_1)^2 \approx \frac{9}{4}U^2\sin^2\Theta - 3U\sum_n A_n\left[\frac{n(n+1)}{\sqrt{(2n+1)(2n+3)}}Y_{n+1}\right.$$

$$-\frac{n(n+1)}{\sqrt{(2n+1)(2n-1)}}Y_{n-1}\Big]\exp(St).$$

The correction to the Laplace pressure $P_\sigma(\xi)$ associated with the nonsphericity of the drop is given by [10]

$$P_\sigma(\xi) = \sum_n (n-1)(n+2)Z_nY_n(\mu)\exp(St).$$

An expression for the pressure $P_E(\xi)$ the electric field exerts on the surface perturbed by the capillary wave motion is derived in the Appendix. Substituting the expressions for $P_E(\xi)$ and $P_\sigma(\xi)$ into (9)–(11), multiplying (9)–(11) by $Y_m(\mu)$, and integrating with respect to the angle Θ we come to a set of homogeneous equations for the amplitudes A_n, B_n, and Z_n:

$$SZ_n - \frac{3}{2}U(\alpha_nZ_{n-1} - \beta_nZ_{n+1}) + (n+1)A_n = 0, \quad (12)$$

$$SZ_n = nB_n, \quad (13)$$

$$-\rho SA_n + SB_n + 3\rho U[\alpha_nA_{n-1} - \beta_nA_{n+1}] + P_\sigma - P_E = 0,$$

$$\alpha_n \equiv \frac{n(n-1)}{\sqrt{(2n-1)(2n+1)}}, \quad (14)$$

$$\beta_n = \frac{(n+1)(n+2)}{(2n+1)(2n+3)}.$$

From (12) and (13), it is easy to find that

$$A_n = \frac{1}{(n+1)}\left\{\frac{3}{2}U(\alpha_nZ_{n-1} - \beta_nZ_{n+1}) - SZ_n\right\},$$

$$B_n = \frac{1}{n}SZ_n.$$

Substituting these relationships into expression (14) yields an infinite set of homogeneous algebraic equations for the amplitudes of the capillary vibrations of the drop:

$$[WeK_n - wT_n]Z_{n-2} - [\rho USL_n + \sqrt{wW}F_n]Z_{n-1}$$
$$+ \{\kappa_nS^2 - WeM_n + \gamma_n\}Z_n - [\rho USI_n + \sqrt{wW}X_n]Z_{n+1}$$
$$+ [WeJ_n - wG_n]Z_{n+2} = 0,$$

$$We \equiv \rho U^2, \quad W \equiv \frac{Q^2}{4\pi\varepsilon}, \quad w = \frac{\varepsilon E^2}{4\pi},$$

$$K_n \equiv \frac{9\alpha_n\alpha_{n-1}}{2n}, \quad L_n \equiv \frac{(9n+6)\alpha_n}{2n(n+1)},$$

$$T_n \equiv 9\frac{n(n-1)(n-2)}{\sqrt{(2n-3)(2n-1)^2(2n+1)}},$$

$$F_n \equiv 3\frac{n(2n-3)}{\sqrt{(2n-1)(2n+1)}}, \quad \kappa_n \equiv \frac{\rho}{(n+1)} + \frac{1}{n}, \quad (15)$$

$$\gamma_n \equiv (n-1)[n+2-W] - w\lambda_n,$$

$$M_n \equiv \frac{9\alpha_n\beta_{n-1}}{2n} + \frac{9\beta_n\alpha_{n+1}}{2(n+2)},$$

$$I_n \equiv \frac{(9n+12)\beta_n}{2(n+1)(n+2)},$$

$$\lambda_n \equiv 9\left(\frac{n^3}{(2n-1)(2n+1)}\right.$$
$$\left. + \frac{(n+1)^2(n+2)}{(2n+1)(2n+3)} - \frac{4n(n+1)-2}{(2n-1)(2n+3)}\right),$$

$$X_n \equiv 3\frac{(n+1)(2n-1)}{\sqrt{(2n+1)(2n+3)}}, \quad J_n \equiv \frac{9\beta_n\beta_{n+1}}{2(n+2)},$$

$$G_n \equiv 9\frac{n(n+1)(n+2)}{\sqrt{(2n+1)(2n+3)^2(2n+5)}}.$$

Here, the variables *We*, *W*, and *w* are the Weber, Rayleigh, and Taylor parameters, respectively.

The necessary and sufficient condition for set (15) to have a solution is that the determinant composed of the coefficients at the amplitudes Z_n equals zero:

$$\begin{vmatrix} \kappa_2S^2 - WeM_2 + \gamma_2 & \rho USI_2 - (wW)^{1/2}X_2 & WeJ_2 - wG_2 & 0\ldots \\ -\rho USL_3 - (wW)^{1/2}F_3 & \kappa_3S^2 - WeM_3 + \gamma_3 & \rho USI_3 - (wW)^{1/2}X_3 & WeJ_3 - wG_3\ldots \\ WeK_4 - wT_4 & -\rho USL_4 - (wW)^{1/2}F_4 & \kappa_4S^2 - WeM_4 + \gamma_4 & \rho USI_4 - (wW)^{1/2}X_4 \\ 0 & WeK_5 - wT_5 & -\rho USL_5 - (wW)^{1/2}F_5 & \kappa_5S^2 - WeM_5 + \gamma_5 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{vmatrix} = 0. \quad (16)$$

Relation (16) is a dispersion relation defining the capillary oscillation spectrum of the drop as a function of dimensionless physical parameters *W*, *w*, and *We*. When these values are varied, the spectrum changes

and some of the solutions S_n^2 may become negative at certain values of *W*, *w*, and *We*. If this condition is fulfilled, the amplitudes of corresponding capillary oscillations grow exponentially with time; in other words, the drop becomes unstable and disintegrates by the law derived in [11]. The condition for the emergence of zero solutions to the dispersion relation is the vanishing of the free term:

$$\begin{vmatrix} \gamma_2 - WeM_2 & -(wW)^{1/2}X_2 & WeJ_2 - wG_2 & 0 & 0 & \dots \\ -(wW)^{1/2}F_3 & \gamma_3 - WeM_3 & -(wW)^{1/2}X_3 & WeJ_3 - wG_3 & 0 & \dots \\ WeK_4 - wT_4 & -(wW)^{1/2}F_4 & \gamma_4 - WeM_4 & -(wW)^{1/2}X_4 & WeJ_4 - wG_4 & \dots \\ 0 & WeK_5 - wT_5 & -(wW)^{1/2}F_5 & \gamma_5 - WeM_5 & -(wW)^{1/2}X_5 & \dots \\ 0 & 0 & WeK_6 - wT_6 & -(wW)^{1/2}F_6 & \gamma_6 - WeM_6 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{vmatrix} = 0. \quad (17)$$

Set (15) is infinite, and determinants (16) and (17) are of an infinitely high order. Therefore, it is reasonable to seek critical conditions for the instability of capillary vibrations (that is, a critical relationship between *W*, *w*, and *We*) by the method of successive approximations, considering two, three, etc. equations of set (15). However, some conclusions concerning solutions to Eq. (17) can be drawn at once.

If the interaction between the modes is neglected, critical conditions for *n*th mode stability that are found by equating the *n*th diagonal component of determinant (17) to zero have the simple analytical form

$$(n-1)(n+2) - WeM_n - (n-1)W - w\lambda_n = 0. \quad (18)$$

Figure 1 shows the critical dependence of the Rayleigh parameter $W = W(We, w)$ on the Weber and Taylor parameters *We* and *w* calculated with (18) for the fundamental mode ($n = 2$). Three numbers corresponding to the coordinates of a point in the space of the variables *W*, *w*, and *We* specify the critical values of the parameters *W*, *w*, and *We*, at which the fundamental mode becomes unstable. From (18) and in Fig. 1, it is seen that the pairwise dependences of the parameters on each other are linear. If two of the three parameters equal zero, the critical values of the third one are as follows: $We = 0.97$, $W = 4$, and $w = 0.86$.

Figure 2 displays the dependence of the Rayleigh parameter *W* on the Taylor parameter *w* for the fundamental mode ($n = 2$) and various *U*. It was calculated with (17) and the 6×6-order determinant. Hereafter in the figures, the ratio of the densities is $\rho = 10^{-3}$. The dependence $W(w)$ in Fig. 2 allows for mode interaction and, unlike that in Fig. 1, is nonlinear. Moreover, if the Rayleigh parameter is zero, the critical Taylor parameter is smaller than in the linear case: $w \cong 0.63$. As follows from Figs. 1 and 2, the critical conditions for the instability of the charged surface of the drop depend appreciably on its velocity, as expected in [12].

Figure 3 plots the real and imaginary components of the dimensionless frequency *S* vs. the dimensionless velocity *U* of the environment at subcritical values of the Rayleigh ($W = 1$) and Taylor ($w = 0.1$) parameters.

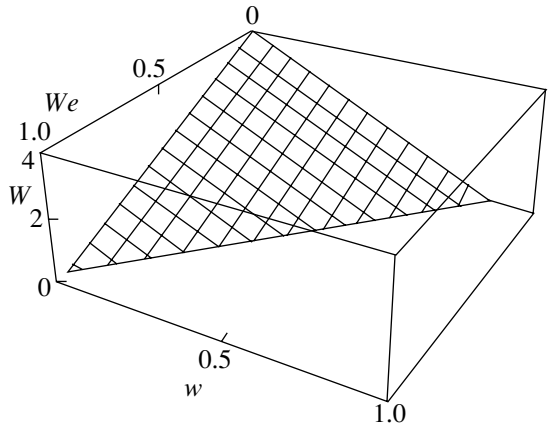


Fig. 1. Relationships between the critical values of the Rayleigh (*W*), Weber (*We*), and Taylor (*w*) parameters for the fundamental (second) mode.

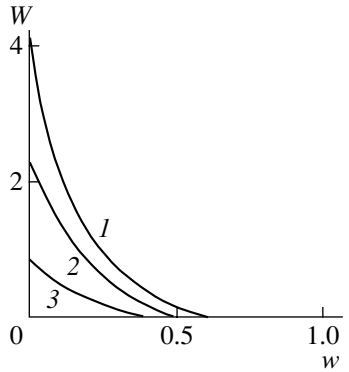


Fig. 2. Critical Rayleigh parameter *W* vs. dimensionless Taylor parameter *w* for the fundamental mode at the dimensionless velocity $U = 0$ (*1*), 20 (*2*), and 30 (*3*).

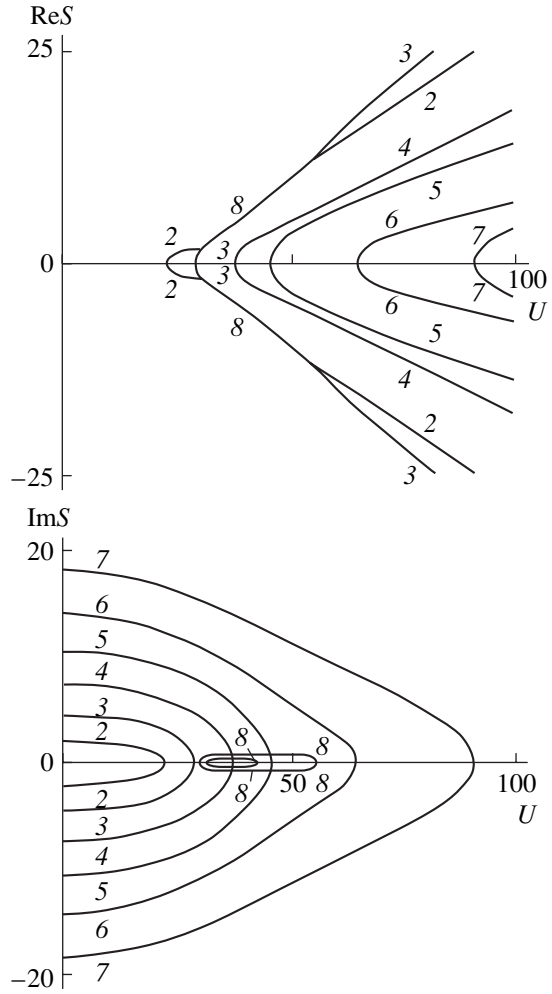


Fig. 3. Real, Re*S*, and imaginary, Im*S*, components of the dimensionless complex frequency *S* of capillary oscillations vs. dimensionless velocity *U* of the flow around the drop for the subcritical value of the Rayleigh parameter $W = 1$ and Taylor parameter $w = 0.1$. Curves *2–7* refer to the associated modes of capillary oscillations of the drop; curve *8*, depicting vibrational instability, is a result of the interaction between the second and third modes.

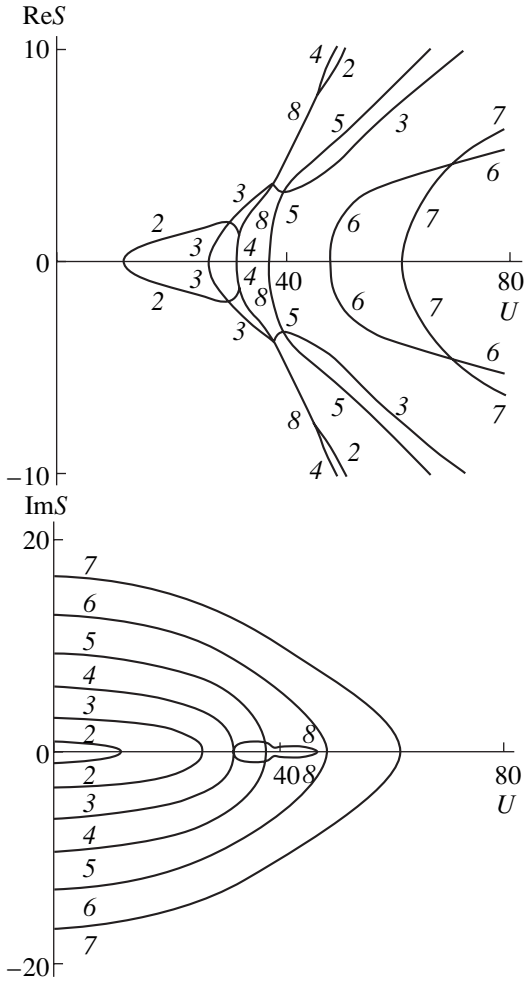


Fig. 4. Real, Re*S*, and imaginary, Im*S*, components of the dimensionless complex frequency *S* of capillary oscillations vs. dimensionless velocity *U* of the flow around the drop of zero charge ($W = 0$) in a strong electric field ($w = 0.6$). Curves *2–7* coincide with the numbers of capillary oscillation modes. Curve *8*, corresponding to the vibrational instability, is a result of the interaction between the second and fourth modes, unlike Fig. 3.

Curves *2–7* correspond to the associated mode numbers. Only the first six equations of set (15) were used in the calculations. The critical dependences for the second and third modes were calculated to three-place accuracy, while those for the sixth and seventh modes are virtually the zeroth approximation. It is of interest that the interaction between the second and third modes causes vibrational solutions *8*. In Fig. 3, the region above the abscissa axis covers unstable solutions; that is, the positive branches of curves *2–8* specify the instability increment for the corresponding mode of the capillary vibrations. Curve *8*, unlike curves *2–7*, has both real and imaginary components. Thus, in a certain range of the velocity *U*, the instability oscillates, which is typical of Kelvin–Helmholtz instability and is associated with the tangential discontinuity of the velocity field at the interface [8, 13].

From Fig. 3, it is seen that, as the velocity increases, the increment of aperiodic instability for the third mode exceeds that for the second mode. This means that if the velocities of the flow around the drop are sufficiently high, the ellipsoidal drop transforms into a parachute-like form. This must show up in the mechanism of drop disintegration in the course of instability development [1]. The vibrational instability of a free-falling drop has been also observed in experiments [1].

It should be noted that the critical velocities for the sixth and seventh modes were calculated at weak mode interaction, because only the first six equations of set (15) were used. Therefore, the critical dependences for these modes contain large errors. This explains the

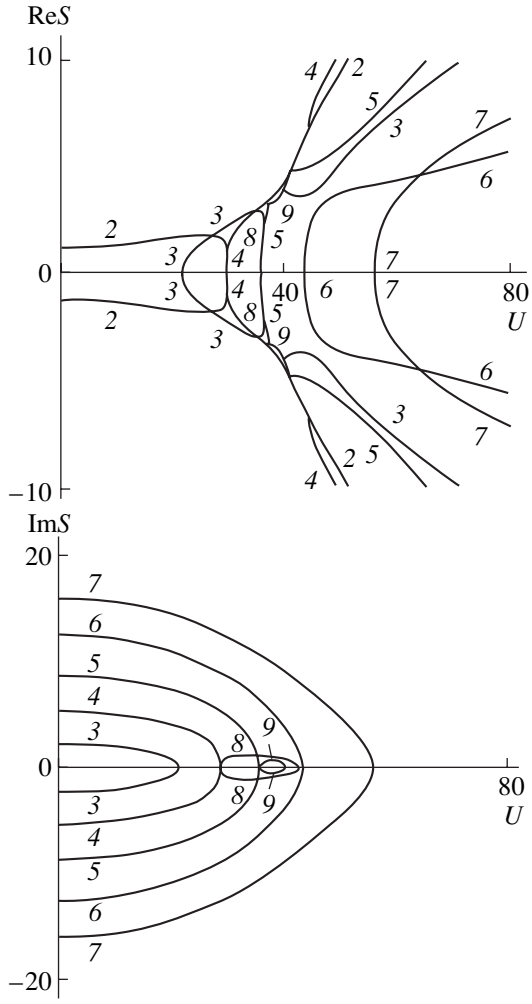


Fig. 5. Real, Re*S*, and imaginary, Im*S*, components of the dimensionless complex frequency *S* of capillary oscillations vs. dimensionless velocity *U* of the flow around the drop at $W = 0$ and a supercritical (for the fundamental mode) value of the Taylor parameter $w = 0.7$. The numbering of the curves is the same as in Fig. 4. Curve *9* reflects the interaction between the third and fifth modes.

qualitative distinction of the dependences Im*S*(*U*) for the sixth and seventh modes from those for the lower modes.

As the Taylor parameter *w* approaches its critical value $w = 0.63$, the shape of the curves calculated and their interaction become qualitatively different in comparison with the case $w \longrightarrow 0$. While the Rayleigh parameter tends to its critical value, the run of the curves remains practically unchanged (they merely shift to the left along the abscissa axis (see [5] and figures therein)). In our case, however, the pattern changes radically. Figure 4 shows dependences similar to those in Fig. 3 but obtained at $W = 0$ and $w = 0.6$. Here, curve *8* is the result of interaction between the second and fourth modes and not between the second and third as in Fig. 3. The vibrational instability described by curve *8* occurs in a rather narrow velocity range, and as the velocity falls outside this range, the fourth mode becomes unstable. Supposedly, this is associated with the fact that the ambient pressure tends to shape the drop into a spheroid flattened along the vector **U** (or $\mathbf{E}_0$), while the electric field pressure extends the spheroidal drop in the same direction. Eventually, the drop will assume a shape close to that specified by the fourth-order polynomial $P_4(\mu)$, and the fourth mode will become unstable.

Figure 5 demonstrates the dependences Re*S*(*U*) and Im*S*(*U*) calculated for the fundamental mode at a slightly supercritical field $\mathbf{E}_0$ ($w = 0.7$ and $W = 0$). It is seen that the behavior of the curves changes only slightly except that the second mode is initially unstable and vibrational branch *9* appears as the result of the interaction between the third and fifth modes.

CONCLUSIONS

Critical conditions for drop instability against the induced charge loosen appreciably in the presence of the self-charge of the drop and a gas flowing around the drop. This results from the superposition of two types of instability: (i) instability of the free surface of the drop against a tangential discontinuity in the velocity field and (ii) instability against the self-charge and the charge induced by an external field. Depending on the ratio between the densities of the drop and the environment, the charge of the drop, the electric field strength, and the velocity of the flow around the drop, the instability can be either aperiodic or vibrational. At low flow velocities, the aperiodic instability shapes the drop into an oblong spheroid and the excess charge is removed by the ejection of finely dispersed heavily charged droplets [11]; at high flow velocities, the drop takes the form of a parachute (at low strengths of the electric field) or becomes cross-shaped (at high strengths of the electric field), disintegrating into many fine droplets and several large drops.

APPENDIX

Pressure Exerted by an External Electric Field and the Field of the Self-Charge of the Drop on the Spherical Surface of a Charged Drop Disturbed by Capillary Wave Motion

Let a perfectly conducting drop carrying a charge *Q* be in a dielectric medium under a uniform electrostatic field $\mathbf{E}_0$.

In the spherical coordinate system with the origin at the drop's center of mass, the pressure of the electric field on the conductor–dielectric interface disturbed by capillary motion is given as follows:

$$r = 1 + \xi: P_E = \frac{\varepsilon}{8\pi}E^2, \quad \mathbf{E} = -\nabla\Phi.$$

Here, Φ is the potential of the total electric field in the dielectric: $\Phi = \Phi_q + \Phi_e = \Phi_0 + \delta\Phi$, where $\Phi_0 = \Phi_q^0 + \Phi_e^0$ is the potential of the electric field without disturbance, and $\delta\Phi = \delta\Phi_q + \delta\Phi_e$ is a correction to the potential Φ_0 due to a weak thermal disturbance $|\xi| \ll R$ of the spherical surface. The correction has the same order of smallness as the disturbance ($|\delta\Phi| \sim |\xi|\backslash R$). The subscripts q and e denote the partial potentials due to the self-charge Q and to the external electrostatic field $\mathbf{E}_0$, respectively. Thus,

$$r = 1 + \xi: P_E = \frac{\varepsilon}{8\pi}(\nabla\Phi_0 + \nabla(\delta\Phi))^2.$$

We expand this expression into a series accurate to terms linear in $|\xi|\backslash R$ and $\delta\Phi$:

$$P_E = \left[\frac{\varepsilon}{8\pi}(\nabla\Phi_0)^2 + \frac{\varepsilon}{8\pi}2\nabla\Phi_0\nabla(\delta\Phi)\right]_{r=1+\xi}$$

$$\approx \left[\frac{\varepsilon}{8\pi}(\nabla\Phi_0)^2 + \frac{\varepsilon}{8\pi}\frac{\partial(\nabla\Phi_0)^2}{\partial r}\xi + \frac{\varepsilon}{4\pi}(\nabla\Phi_0)(\nabla(\delta\Phi))\right]_{r=1}.$$

Since the electric field on the conductor surface is normal to the surface, we obtain

$$r = 1: P_E = \frac{\varepsilon}{8\pi}\left(\frac{\partial\Phi_0}{\partial r}\right)^2 + \frac{1}{8\pi}\left[\frac{\partial}{\partial r}\left(\frac{\partial\Phi_0}{\partial r}\right)^2\right]\xi + \frac{1}{4\pi}\frac{\partial\Phi_0}{\partial r}\frac{\partial(\delta\Phi)}{\partial r}. \quad \text{(A.1)}$$

The contributions from the charge and external field to the total potential Φ are independent of each other. We therefore divide the problem into two subproblems and will seek the partial potentials separately.

The potential Φ_e is a solution to the subproblem

$$\Delta\Phi_e = 0,$$

$$r \longrightarrow \infty: \Phi_e \longrightarrow -E_0 r\cos\Theta,$$

$$r \longrightarrow 1 + \xi: \Phi_e = \text{const}.$$

For the zeroth, Φ_e^0, and first, $\delta\Phi_e$, orders of smallness, we have the boundary-value problems

$$\Delta\Phi_e^0 = 0, \quad \Delta(\delta\Phi_e) = 0, \quad \text{(A.2)}$$

$$r \longrightarrow \infty: \Phi_e^0 \longrightarrow -E_0 r\cos\Theta, \quad \delta\Phi_e \longrightarrow 0, \quad \text{(A.3)}$$

$$r = 1: \Phi_e^0 = 0, \quad \frac{\partial\Phi_e^0}{\partial r}\xi + \delta\Phi_e = 0. \quad \text{(A.4)}$$

Solving the first equation (A.2) in the zeroth approximation subject to boundary conditions (A.3) and (A.4) yields

$$\Phi_e^0 = -E_0 r\cos\Theta(1 - r^{-3}). \quad \text{(A.5)}$$

A solution to the boundary-value problem in the first approximation for $\delta\Phi_e$ is natural to seek in the form

$$\delta\Phi_e = \sum_n B_n(t)r^{-(n-1)}.$$

Taking into account boundary conditions (A.3) and (A.4), expansion (7) for $\xi(\Theta, t)$, and the recurrent relationship for spherical functions [9]

$$\cos\Theta Y_n(\mu) = \frac{(n+1)}{\sqrt{(2n+1)(2n+3)}}Y_{n+1}(\mu) + \frac{n}{\sqrt{(2n+1)(2n-1)}}Y_{n-1}(\mu), \quad \text{(A.6)}$$

we find

$$\delta\Phi_e = 3E_0\sum_{n=2}^{\infty}\frac{1}{r^{n+1}}\left[\frac{(n+1)Z_{n+1}}{\sqrt{(2n+1)(2n+3)}} + \frac{(n+1)Z_{n-1}}{\sqrt{(2n+1)(2n+3)}}\right]Y_n(\mu). \quad \text{(A.7)}$$

The potential Φ_q is a solution to the boundary-value problem

$$\Delta\Phi_q = 0,$$

$$r \longrightarrow \infty: \Phi_q \longrightarrow 0,$$

$$r = 1 + \xi: \Phi_q = \text{const}.$$

For the zeroth, Φ_q^0, and first, $\delta\Phi_q$, orders of smallness, we arrive at the boundary-value problems

$$\Delta\Phi_q^0 = 0, \quad \Delta(\delta\Phi_q) = 0, \quad \text{(A.8)}$$

$$r \longrightarrow \infty: \Phi_q^0 \longrightarrow 0, \quad \delta\Phi_q \longrightarrow 0, \quad \text{(A.9)}$$

$$r = 1: \Phi_q^0 = Q, \quad \delta\Phi_q = -\frac{\partial\Phi_0}{\partial r}\xi. \quad \text{(A.10)}$$

A solution to the first equation (A.8) in the zeroth approximation subject to boundary conditions (A.9) and (A.10) has the expectable form

$$\Phi_q^0 = \frac{Q}{\varepsilon r}. \quad \text{(A.11)}$$

A solution for $\delta\Phi_q$ will be sought in the form

$$\delta\Phi_q = \sum_n D_n(t)r^{-(n-1)}.$$

Taking into account boundary conditions (A.9) and (A.10) in view of expansion (7) for $\xi(\Theta, t)$, we obtain

$$\delta\Phi_q = \frac{Q}{\varepsilon}\sum_n\frac{1}{r^{n+1}}Z_nY_n(\mu). \quad \text{(A.12)}$$

Substituting expressions (A.5), (A.7), (A.11), and (A.12) into expression (A.1) for the pressure and also using recurrent relations (A.6) and the expression [9]

$$(2n-1)(2n+3)\cos^2\Theta Y_n(\mu) = (2n-1)$$
$$\times\frac{(n+1)(n+2)}{\sqrt{(2n+1)(2n+5)}}Y_{n+2}(\mu) + [2l(l+1)-1]Y_n(\mu)$$
$$+(2n+3)\frac{n(n-1)}{\sqrt{(2n+1)(2n-3)}}Y_{n-2}(\mu),$$

we arrive at

$$P_E = \frac{W}{2} + 3\sqrt{wW}\cos\Theta + 18\cos^2\Theta$$
$$+\sum_n Y_n(\mu)\Bigg\{9w\frac{n(n-1)(n-2)}{\sqrt{(2n-3)(2n-1)^2(2n+1)}}Z_{n-2}$$
$$+3\sqrt{wW}\frac{n(2n-3)}{\sqrt{(2n-1)(2n+1)}}Z_{n-1}$$
$$+\Bigg[W(n-1)+9w\Bigg(\frac{n^3}{(2n-1)(2n+1)} \qquad \text{(A.13)}$$
$$+\frac{(n+1)^2(n-2)}{(2n+1)(2n+3)}-\frac{4n(n+1)-2}{(2n-1)(2n+3)}\Bigg)\Bigg]Z_n$$
$$+3\sqrt{wW}\frac{(n+1)(2n-1)}{\sqrt{(2n+1)(2n+3)}}Z_{n+1}$$
$$+9w\frac{n(n+1)(n+2)}{\sqrt{(2n+1)(2n+3)^2(2n+5)}}Z_{n+2}\Bigg\}.$$

REFERENCES

1. A. L. Gonor and V. Ya. Rivkind, Itogi Nauki Tekh., Ser. Mekh. Zhidk. Gaza **17**, 98 (1982).
2. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).
3. D. F. Belonozhko and A. I. Grigor'ev, Elektrokhim. Obrab. Met., No. 4, 17 (2000).
4. V. A. Dyachuk and V. M. Muchnik, Dokl. Akad. Nauk SSSR **248**, 60 (1979).
5. A. I. Grigor'ev, V. A. Koromyslov, and S. O. Shiryaeva, Zh. Tekh. Fiz. **69** (5), 7 (1999) [Tech. Phys. **44**, 486 (1999)].
6. A. I. Grigor'ev, V. A. Koromyslov, and S. O. Shiryaeva, Zh. Tekh. Fiz. **70** (7), 26 (2000) [Tech. Phys. **45**, 840 (2000)].
7. G. Taylor, Proc. R. Soc. London, Ser. A **280**, 383 (1964).
8. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
9. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
10. S. O. Shiryaeva, A. É. Lazaryants, *et al.*, Preprint No. 27, IMRAN (Yaroslavl, 1994).
11. A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **61** (3), 19 (1991) [Sov. Phys. Tech. Phys. **36**, 258 (1991)].
12. A. I. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **28** (5), 12 (2002) [Tech. Phys. Lett. **28**, 176 (2002)].
13. A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **66** (2), 23 (1996) [Tech. Phys. **41**, 124 (1996)].

Translated by N. Mende

Technical Physics, Vol. 47, No. 9, 2002, pp. 1098–1105. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 29–37.

GAS DISCHARGES, PLASMA

Deuterium Liner and Multiparametric Studies of the Formation of an Inverse Z-Pinch

Vit. M. Bystritskii[1], Vyach. M. Bystritsky[2], J. Wozniak[3], V. M. Grebenyuk[2], E. Gula[4], G. N. Dudkin[5], G. A. Mesyats[6], B. A. Nechaev[5], V. N. Padalko[5], S. S. Parzhitski[2], F. M. Pen'kov[2], N. A. Ratakhin[7], S. A. Sorokin[7], and V. A. Stolupin[2]

[1]*University of California, Irvine, CA 92697-4575 USA*
[2]*Joint Institute of Nuclear Research, Dubna, 141980 Russia*
[3]*Faculty of Physics and Nuclear Technology, Krakov, 30059 Poland*
[4]*Institute of Nuclear Physics, Krakov, 31342 Poland*
[5]*Research Institute of Nuclear Physics at TPU, Tomsk, 634055 Russia*
[6]*Institute of Electrophysics, Yekaterinburg, 620219 Russia*
[7]*Institute of High-Current Electronics, Tomsk, 634055 Russia*
e-mail: bystvm@nusun.jinr.ru
Received February 1, 2002

Abstract—A method and results of measurements are presented of the ion energy distribution in a deuterium liner accelerated in the inverse Z-pinch, in which the plasma is accelerated electrodynamically from the liner axis. Knowledge of the deuteron energy distribution is of primary importance for the correct interpretation of the experimental results from the study of the *dd*-reaction in the range of infralow energies with the use of a liner plasma. Experiments were carried out in a high-current pulsed accelerator (I = 950 kA, τ = 80 ns) at the Institute of High-Current Electronics of the Siberian Division of the Russian Academy of Sciences (Tomsk, Russia). In the initial state, the liner is a supersonic hollow deuterium jet 32 mm in diameter and 20 mm in length. The liner parameters were measured with the help of optical detectors of H_α and H_β deuterium lines and magnetic probes arranged in a radial direction (along the direction the liner expansion). In addition, scintillation spectrometers and BF_3 counters were used to measure the intensity of the neutron flux produced in the $d + d \longrightarrow {}^3\mathrm{He} + n$ reaction. The results obtained by simultaneously analyzing the data from magnetic probes, optical detectors, and neutron detectors point to the possibility of using a rather simple method for measuring the parameters of the liner accelerated up to energies of 3–6 keV. *© 2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Interest in the study of strong interactions between light nuclei in the range of infralow energies (~1 keV) is motivated not only by the possibility of verifying the fundamental symmetries, but also by the necessity of solving a number of astrophysical problems [1]. Until recently, experimental data on the cross sections for reactions between light nuclei in the range of infralow collision energies (from several electronvolts to several kiloelectronvolts) have almost completely been lacking. This is explained by both the small values of the cross sections (10^{-43}–10^{-32} cm^2) and the very low intensities of charged-particle beams.

Experimental studies of nuclear reactions in this energy range have been stimulated by the proposition of a new method for studying such reactions [1–4]. This method is based on the use of high-intensity converging ion flows generated in the process of the liner plasma implosion (the formation of a straight Z-pinch). In those experiments, the liner was a 32-mm-diameter supersonic hollow cylindrical deuterium jet injected through a Laval nozzle into a vacuum, and the target was either a hollow deuterium jet with a radius smaller than the liner radius or a CD_2 rod positioned along the liner axis. With this method, a number of experiments [5–9] were performed to study the nuclear reaction

$$d + d \longrightarrow {}^3\mathrm{He} + n \qquad (1)$$

in the deuteron collision energy range 0.05–2.3 keV. The effective cross sections for the *dd*-reaction and the astrophysical *S*-factor in the above energy range were estimated. The results obtained testify that, on the one hand, the proposed method deserves attention and can be efficiently used to realize the program of studying nuclear reactions in the astrophysical energy range and, on the other, this method can be regarded as a basis for nuclear-physics diagnostics of the plasma processes. Note that, when this method is used to obtain precise data on the parameters of nuclear reactions (cross sections and astrophysical *S*-factors), it is required that the energy distribution of the ions accelerated in the liner, as well as the Coulomb deceleration of the liner ions in the course of interaction with the target, be known to a sufficiently high accuracy [8–10]. This requirement is

1063-7842/02/4709-1098$22.00

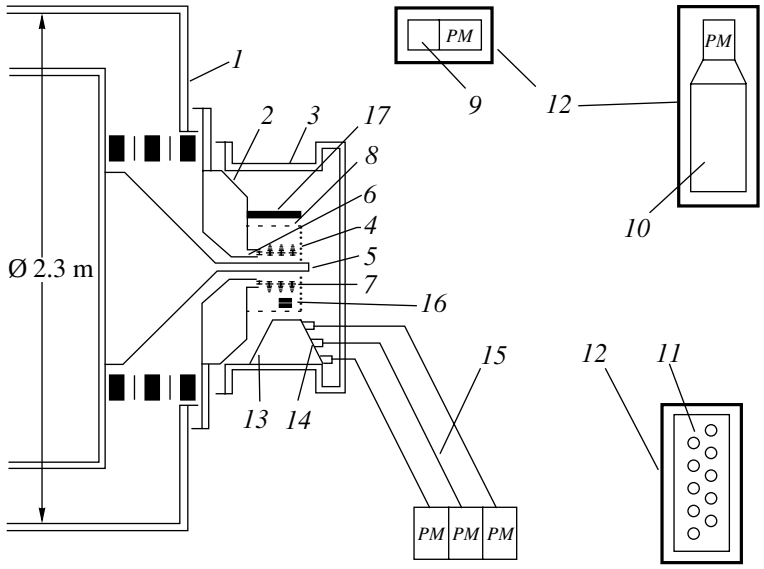


Fig. 1. Schematic of the experimental setup: (*1*) high-current generator, (*2*) accelerator load unit, (*3*) measuring chamber, (*4*) grid cathode, (*5*) reverse-current conductor, (*6*) supersonic Laval nozzle, (*7*) liner, (*8*) current-interception rod, (*9*) scintillation detector *D*1, (*10*) scintillation detector *D*2, (*11*) thermal neutron detector *D*3, (*12*) Pb protective shield, (*13*) light cone, (*14*) collimators, (*15*) fibers, (*16*) *dB*/*dt* magnetic probes, and (*17*) CD_2 target. The electromagnetic valve is not shown.

imposed because the cross section for the nuclear reactions under study show a sharp dependence on the collision energy. In the range of infralow energies, this dependence is exponential.[1] We also mention some disadvantages of the proposed method for studying nuclear reactions. Because of the onset of various types of instabilities during the liner implosion (e.g., MHD instabilities), high-energy (20–300 keV) ions can be generated in the liner at a level of 10^{-10}–10^{-12} of the total number of accelerated ions. This, in turn, promotes nuclear reactions (the background process) and, as a consequence, considerably increases the measured yield of their products. In the straight Z-pinch scheme, it is difficult to resolve in time the processes of electrodynamic ion acceleration in the liner and the processes of the liner interaction with the target; for this reason, when analyzing the experimental data, it becomes problematic distinguishing the background events and the events under study. Another disadvantage of the method proposed is the limitation of the maximum attainable energy of the accelerated ions at a level of <3 keV; otherwise, the contribution of the background processes to the yield of the nuclear reaction under study increases. To decrease the contribution of the background processes and to increase the maximum attainable energy of the accelerated liner ions up to ~15 keV, we proposed a method based on the formation of an inverse Z-pinch [11]. The production of high-energy (6–15 keV) accelerated ions is required to compare the parameters of nuclear reactions measured with use of conventional and plasma accelerators.[2]

In the inverse Z-pinch, the liner plasma is accelerated from the liner axis in the radial direction. This makes it possible (i) to resolve in time the processes of the liner electrodynamic acceleration and the processes of the liner interaction with the target, (ii) to reduce the intensity of the plasma flow onto the target, and (iii) to substantially simplify the problem of measuring the average density and the ion velocity distribution in the liner as it approaches the target. The results of studying the formation of an inverse Z-pinch [11] testify that there is the possibility of time resolving "useful" and background events; consequently, this method is applicable to studying nuclear reactions in the infralow energy range. However, even in this case, to obtain information about the ion energy distribution in the liner by the inverse-Z-pinch method is a rather complicated problem that requires the use of various, sufficiently simple and highly reliable methods. This study is aimed at solving this problem.

EXPERIMENT

Experiments were carried out at the Institute of High-Current Electronics (Siberian Division, Russian Academy of Sciences) with the use of an SGM pulsed high-current accelerator (the liner current was $I = 950$ kA and the high-voltage pulse duration was $\tau = 80$ ns) [14]. Figure 1 shows a schematic of the experimental setup

[1] Information about the main parameters of the reactions under study is obtained by using the parametrization of the energy dependence of the reaction cross section, which can be represented as the product of the astrophysical *S*-factor and the barrier factor related to the Coulomb repulsion of nuclei.

[2] The lowest energy for which the cross section for the *dd*-reaction with the production of neutrons was measured by using conventional accelerators is 6 keV in the center-of-mass frame of reference of the colliding deuterons [12, 13].

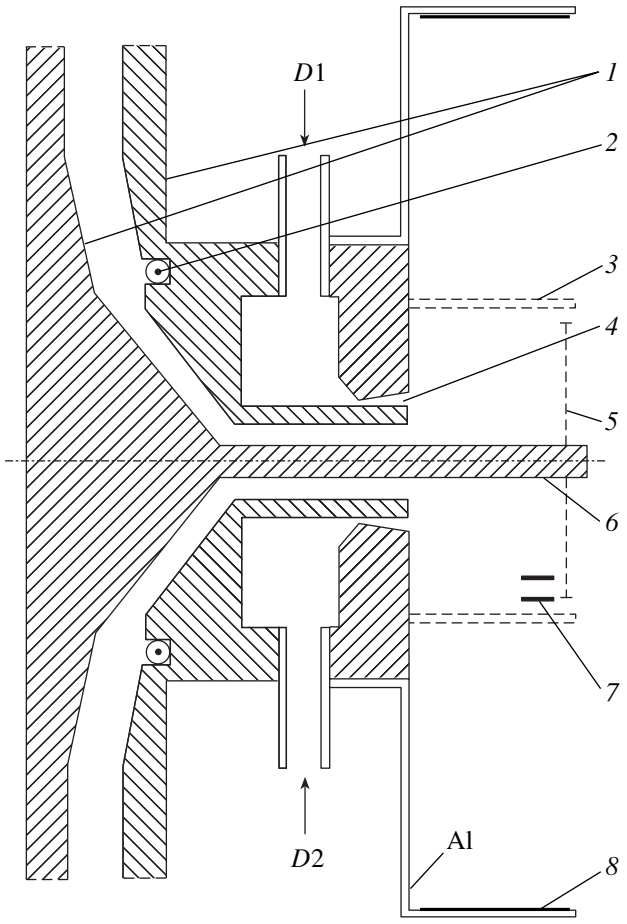


Fig. 2. Sketch of the load used to form an inverse Z-pinch and the positions of the optical detectors: (*1*) electrodes, (*2*) Rogowski loop, (*3*) current-interception rods, (*4*) Laval nozzle, (*5*) grid cathode, (*6*) reverse-current conductor, (*7*) magnetic probes, and (*8*) target.

consisting of a high-current generator, a load unit (a sketch of the load unit is presented in Fig. 2), and a detecting and diagnostic facility.

1. Method for Producing an Inverse Z-pinch

An original deuterium liner is produced with the help of a Laval nozzle (*6*) (Fig. 1) and a fast electromagnetic valve. The nozzle is placed on the grounded electrode (*1*) of the diode of the high-current generator. The liner mean radius amounts to 15 mm. A current-interception structure (*8*) (CIS) is placed on the path of radial expansion of the plasma shell; this structure has the form of a 45-mm-radius squirrel cage made from rods 1 mm in diameter. The CIS position radius determines the liner-acceleration length. The current flowing through the liner was measured by a Rogowski loop (Fig. 2). The liner mass was assessed by using a zero-dimensional model and information about the time dependence of the current flowing through the liner, along with the start times of the magnetic probe signals (dB/dt), which correspond to the times when the current-carrying shell passes through the probes.

The process of liner acceleration is monitored with two dB/dt probes located at radii of 23 and 34 mm.

2. Method for Measuring the Deuteron Energy Distribution in the Liner

The method proposed is based on the detection of optical radiation propagating in the radial direction from the axis of the deuterium liner. The liner radiation is generated in the following processes: the charge exchange of deuterium ions with molecules (atoms) of the residual gas in the measuring chamber of the accelerator and the subsequent emission of the deuterium H_α 656.5-nm and H_β 486-nm lines and bremsstrahlung. To separate the Balmer lines, we used glass filters in the blue-green and red spectral regions. Three optical detectors (*LD*1, *LD*2, and *LD*3) were arranged behind the CIS in the radial direction (in the direction of the liner expansion) (Fig. 1). The distance from the CIS to the detector *LD*1, as well as the *LD*1–*LD*2 and *LD*2–*LD*3 distances, was equal to 50 mm. Each optical detector consisted of a collimator, quartz fiber, and photomultiplier. The diameter and length of the collimator amounted to 4 and 14 mm, respectively. The diameter of the quartz fiber was 200 μm, and its length was 7 m. To protect the optical detectors from the background light emitted during the breakdown of the gas jet injected into the diode, the collimators were enclosed in an oval truncated cone made from an opaque dielectric. The cone had the following dimensions: the inlet diameter was 20 mm, the height was 220 mm, the longer axis of the cone base was 110 mm, and the shorter axis was 40 mm. The method for measuring the ion energy distribution in the liner is based on the relationships between the start times of the signals from the optical detectors located at fixed distances from the CIS, as well as the durations of these signals. The signal duration is determined by the light-pulse duration, which is a function of the distance from the current-interception rods. In this approach, the liner ion energy distribution can be represented in a simplified form as

$$\Delta t[\text{ns}] \approx 16.09 L (1/E)^{1/2} (\Delta E/E), \qquad (2)$$

where Δt is the full widths at half-maximum (FWHM) of the light pulse from the detector located at the distance L (in cm) from the CIS (it is assumed that, when the liner reaches the CIS, the liner current is intercepted by the CIS and the further expansion of the liner proceeds without acceleration and is, in fact, the free motion of a current-free shell), E is the most probable energy of the liner ions (in keV), and ΔE is the full width of their energy distribution at the distance L from the CIS.

Hence, the broadening of the light-detector signals with increasing distance between the liner and the CIS corresponds to the spread in the energies of the liner

ions. We should mention postulates adopted in the approach described above. First, it is assumed that the spatial extension of optical radiation in the H_α and H_β lines corresponds to the distribution of the liner ions. This postulate is based on the assumption that all the liner ions start accelerating at the same time irrespective of their initial radial position (at the instant when the high-voltage pulse is applied to the gas liner, its radial thickness is ~1–2 cm). Second, it is assumed that, by measuring the intensity of the H_α and H_β lines, we obtain information about the radial distribution of the liner ion density. This assumption is valid if there is thermodynamic equilibrium between the liner ions and the excited neutrals emitting in the H_α and H_β lines. The liner parameters were measured with a TEKTRONIX oscilloscope, and the trigger signal was the current pulse of the high-voltage generator.

3. Bolometric Studies of the Liner Expansion

The purpose of the bolometric measurements was to determine the energy flux density ΔW in the radially expanding plasma shell. In our experiment, this quantity was measured by a foil bolometer. The method of bolometric measurements consists in deducing the energy of the liner ions arriving at the foil from the effects of the heating and associated change in the electrical resistance of the foil.

The bolometric measurement circuit includes a reference-current generator, foil bolometer, and recording channel on the basis of a TEKTRONIX oscilloscope.

The reference-current generator consists of an artificial forming line having a wave impedance of ~2 Ω which is switched to a matched active load ~2 Ω by a fast thyristor.

The foil bolometer with a resistance of ~0.05–0.1 Ω was connected in series to a load resistor. The foil bolometer was a copper foil strip with a length of l = 45 mm, width of 2 mm, and thickness of δ = 17.5 μm. This strip was placed in parallel with the liner axis at a radius of r = 360 mm. The amplitude and duration of the reference current pulse of nearly squared shape were ~60 A and ~10 μs, respectively; this guaranteed the low Joule heating of the foil.

The liner plasma energy flux per 1 cm^2 of the copper foil surface is related to the relative change in the voltage on it, $\Delta U/U$, by the relationship

$$\Delta W[\mathrm{J/cm^2}] = C_p\rho\delta\alpha\Delta U/U = 1.56\Delta U/U, \quad \Delta U/U = \alpha\Delta T, \tag{3}$$

where C_p is the specific heat capacitance of the copper foil (C_p = 0.39 J/gK), α is the temperature coefficient of the foil resistance ($\alpha = 3.9 \times 10^{-3}$ K^{-1}), ρ is the copper mass density (ρ = 8.93 g/cm^3), and ΔT is the change in the temperature of the copper foil after absorbing the incident energy flux.

Formula (3) was derived assuming the linear dependence of ΔW on ΔU, which takes place if C_p and α are constant over the foil temperature interval ΔT. For the given values of C_p and α, formula (3) is valid in the temperature range $\Delta T \le 300$ K, which is precisely the temperature variation range under the conditions of our experiments. For $\Delta T \le 300$ K, the bolometer design described above allowed us to measure plasma energy fluxes of up to ΔW = 1.8 J/cm^2. The accuracy of the bolometer measurements is limited by the accuracy of electrical measurements and an uncertainty in the energy absorption factor of the foil. The possibility of measuring the energy flux of the light ions arriving at the foil (H^+, He^+, and Li^+) is provided by the relatively low reflection coefficient (<20%) of the ions with energies from several hundreds of electronvolts to tens of kiloelectronvolts. The cathode sputtering of the bolometer material only slightly affects the accuracy of measurements, because the sputtering factor is usually less than unity and the mean particle sputtering energy does not exceed several electronvolts.

It should also be noted that the condition of the weak heating of the bolometer foil by the reference current sets limits on the value of the voltage U across the foil (several volts) and the change in this voltage ΔU (~1 V). There are some problems in measuring such signals in the presence of large electromagnetic pickups arising during the operation of the high-current generator. To suppress this type of pickup, the recording oscillograph and the reference-current generator were placed inside a special shielding box. They were supplied with power through a supply-line filter. The signal cables were enclosed into an additional shielding braid. The starts of the high-current generator and reference-current generator were synchronized with the help of a GI-1 6-channel delayed-pulse generator.

As an example, Fig. 3 shows the bolometer signal obtained in one of the shots (curve *1*). The figure also shows the reference signal recorded without switching on the high-current generator (curve *2*). The instant at which spikes appear in the reference signal coincides with the start of the high-current generator. These spikes are associated with both the transient processes occurring in the reference-current generator circuit when the resistance changes rapidly and the partial shunting of the active element by the plasma flow. The ratio $\Delta U/U$ was measured with a time delay $\Delta t > 1$ μs relative to the start of the high-current generator, i.e., after the transient processes in the circuits had come to an end. As is seen in Fig. 3, the relative change in the voltage $\Delta U/U$ varies only slightly over the measurement time interval; thus, we could determine the increase in the foil temperature and, consequently, the energy flux density near the foil position with good accuracy.

In this shot, we have $\Delta U/U$ = 0.26, which corresponds to the energy flux density of the liner plasma $\Delta W = 1.56\Delta U/U$ = 0.4 J/cm^2 and the energy flux

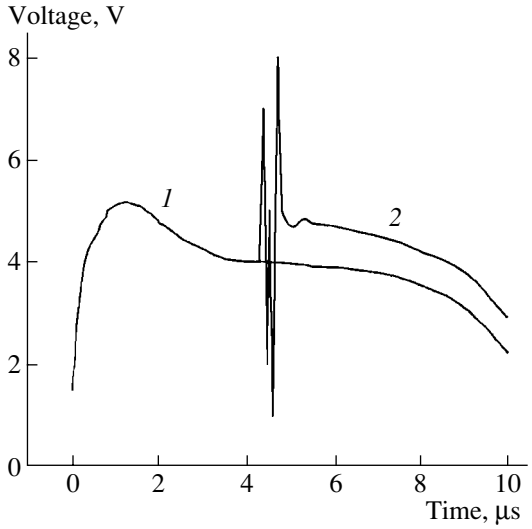


Fig. 3. Bolometer signals: (*1*) reference signal at the output of the bolometer and (*2*) bolometer signal when the high-current generator is switched on.

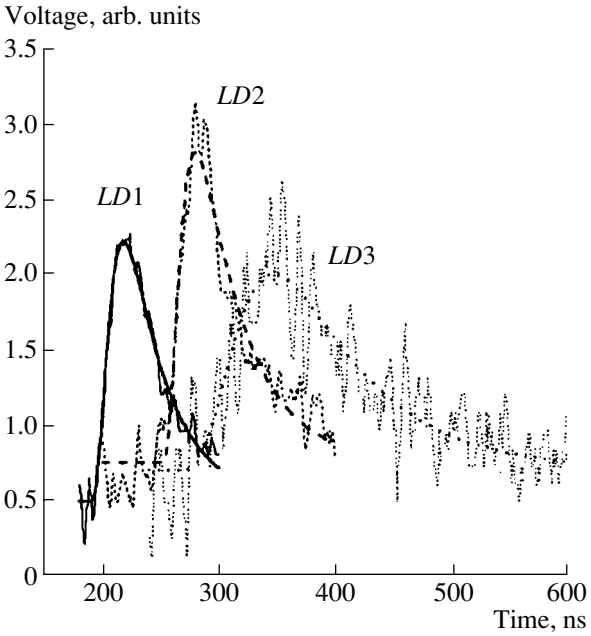


Fig. 4. Signals from optical detectors *LD*1, *LD*2, and *LD*3 obtained in shot no. 3.

absorbed by the target $E = \Delta WS = 400$ J (where $S = 2\pi rl = 1017$ cm^2 is the area of the target enclosing the liner). Zero-dimensional calculations of the liner acceleration dynamics with the use of the data from magnetic probes placed on the liner-acceleration path show that energy spent on the liner acceleration amounts to ~1 kJ (the liner mass is ~8 μg, the length is 20 mm, and the final liner velocity is $\sim 5 \times 10^7$ cm/s). Consequently, the target of length 45 mm placed around the liner at a radius of 360 mm absorbs ~40% of the liner kinetic energy. In the next shot, the bolometer was displaced by 5 cm along the liner axis, its radial coordinate being the same. In this case, the measured energy flux amounted to ~200 J. From the results obtained, it follows that the plasma expands predominantly in the radial direction and its width (along the liner axis) at a radius of 36 cm can be estimated at 10–15 cm. The measured kinetic energy of the plasma flow agrees with that calculated by the zero-dimensional model.

4. Liner Diagnostics with Neutron Detectors

To measure the energy of the *dd* neutrons produced as a consequence of the liner–target interaction or arising due to the onset of instabilities during the liner acceleration, we used the time-of-flight neutron detection technique.

The neutrons (Fig. 1) produced in reaction (1) were detected with two scintillation detectors *D*1 ($d = 5.3$ cm, $l = 5.0$ cm) and *D*2 ($d = 10$ cm, $l = 20$ cm) on the basis of plastic scintillators and with a thermal-neutron detector *D*3 (an assembly of ten BF_3 proportional counters placed in a polyethylene moderator). Detectors *D*1–*D*3 were placed at distances of 246, 410, and 277 cm from the liner axis, respectively. To protect the detectors from high-intensity bremsstrahlung and X radiation, they were surrounded by a 5-cm-thick Pb shield. The neutron detection efficiencies of detectors *D*1–*D*3 amounted to 2×10^{-6}, 4×10^{-6}, and 4.3×10^{-6}, respectively.

ANALYSIS AND DISCUSSION OF RESULTS

1. Optical Detectors

As an example, Fig. 4 shows the optical detector signals in shot no. 3 and their approximations by "pulse" functions. The time shifts between the peaks of the optical detector signals correspond to the average liner velocity. Based on this information, the time dependences of the optical detector signals can be easily recalculated to the ion energy distributions (Fig. 5). In principle, according to the dynamics of a current-free liner, the shifts between the signals from detectors *LD*1 and *LD*2 and those from detectors *LD*2 and *LD*3 should be the same. To verify this inference, the optical detector signals were processed separately for intervals 1–2 and 2–3.

A scenario of the formation of optical radiation may be described as follows. The liner, which has a certain initial ion energy distribution after passing the current-interception rods, expands radially in the residual gas with the density $n \sim 10^{13}$ cm^{-3}. This expansion is accompanied by the charge exchange of the liner ions with the residual gas (predominantly, nitrogen) molecules,

$$D^+ + N_2 \longrightarrow D^* + N_2^+ \longrightarrow D + N_2^+ + h\nu, \qquad (4)$$

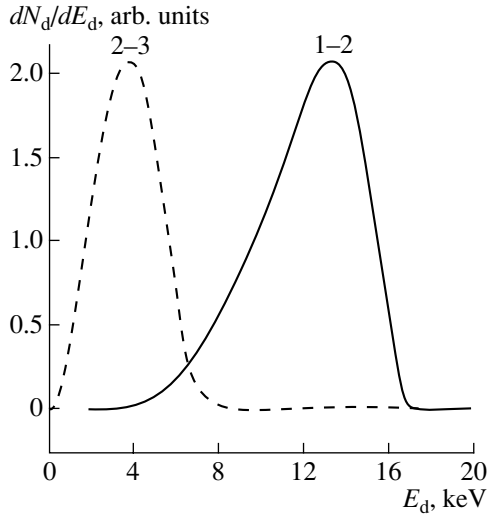


Fig. 5. Energy distributions of the liner deuterons for intervals 1–2 and 2–3 in shot no. 3.

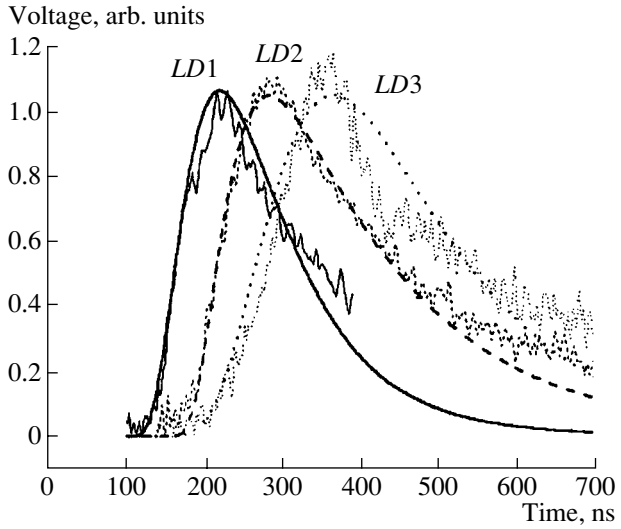


Fig. 6. Signals from optical detectors *LD*1– *LD*3 obtained in shot no. 7.

as well as by bremsstrahlung emission and impact–radiative recombination [15],

$$e^- + D^+ + e^- \longrightarrow D + e^- + h\nu, \quad (5)$$

$$e^- + D^+ \longrightarrow D + h\nu. \quad (6)$$

These processes determine the formation of optical radiation from the liner; they are also responsible for a discrepancy between the values of the liner parameters deduced from an analysis of the data for intervals 1–2 and 2–3 (Table 1). This discrepancy can be explained by the fact that, at distances no longer than 5–10 cm from the CIS, the light intensity [process (4)] is primarily determined by the "fast" ion component of the liner (the fraction of these ions amounts to a few percent of the total number of the liner ions). This follows from the shape of the dependence of the cross section for process (4) on the ion energy and from the fact that the nitrogen atom density in the spatial region through which the fast liner component has passed is substantially lower than its initial value. Fast excited deuterium atoms [see process (3)] emit light for a characteristic time of ~10 ns; then, these neutral atoms can be excited in the processes $e^- + D \longrightarrow D^- + h\nu$ and $e^- + D \longrightarrow D^* + e^-$ [15]. Note that the cross section for these processes ($\sigma \le 10^{-17}$ cm^2) is much less than for process (4).

At distances longer than 10 cm from the CIS, the generation of optical radiation is mainly determined by

Table 1. Results of processing of signals from optical detectors

Shot no.	1	2	3	4	5	6	7
$T_2 - T_1$, ns	56	43	45	59	45	71	70
$T_3 - T_2$, ns	76	96	84	76	70	70	72
$\Delta T_2 - \Delta T_1$, ns	–	11	10	51	16	40	50
$\Delta T_3 - \Delta T_2$, ns	45	77	60	37	28	35	55
V_{21}, 10^7 cm/s	8.9 ± 0.4	11.6 ± 0.9	11.1 ± 0.6	8.4 ± 0.3	11.1 ± 0.6	7.0 ± 0.2	7.1 ± 0.3
E_{21}, keV	8.2 ± 0.7	13.9 ± 1.5	12.8 ± 1.4	7.4 ± 0.6	12.8 ± 1.4	5.1 ± 0.3	5.2 ± 0.4
ΔE_{21}, keV	–	7.1 ± 2.5	5.7 ± 2.0	12.9 ± 0.8	9.1 ± 2.0	5.8 ± 0.5	8.2 ± 0.5
V_{32}, 10^7 cm/s	6.5 ± 0.2	5.2 ± 0.1	5.9 ± 0.2	6.5 ± 0.2	7.1 ± 0.3	7.1 ± 0.3	6.9 ± 0.2
E_{32}, keV	4.5 ± 0.3	2.8 ± 0.1	3.6 ± 0.2	4.5 ± 0.3	5.3 ± 0.4	5.3 ± 0.4	4.9 ± 0.3
ΔE_{32}, keV	3.9 ± 0.4	2.8 ± 0.2	3.9 ± 0.3	3.2 ± 0.4	3.1 ± 0.5	3.9 ± 0.5	4.8 ± 0.4

Note: T_1, T_2, and T_3 are the times corresponding to the amplitudes of the voltage pulses at the outputs of detectors *LD*1, *LD*2, and *LD*3, respectively; ΔT_1, ΔT_2, and ΔT_3 are the FWHMs of the signals; V_{21} and V_{32} are the deuteron velocities deduced from the time shift of the signals from two neighboring detectors; E_{21} and E_{32} are the deuteron energies corresponding to the velocities V_{21} and V_{32}; and ΔE_{21} and ΔE_{32} are the FWHMs of the deuteron energy distribution determined from the broadening of the signals from the two detectors. The accuracy of estimates for T_1, T_2, and T_3 is 2.0 ns, and the accuracy of estimates for ΔT_1, ΔT_2, and ΔT_3 is 3.0 ns. Bars in the columns mean the absence of relevant information.

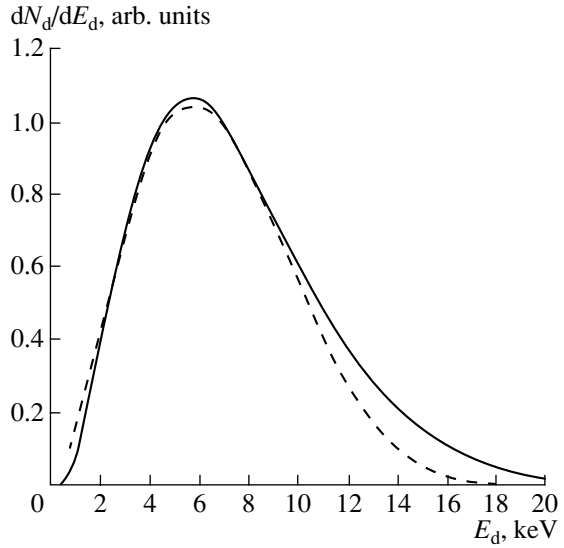


Fig. 7. Energy distributions of the liner deuterons in shot no. 7 at the shifted positions of the optical detectors with respect to the current-interception rods. The shift from the initial position is 5 cm. The solid line corresponds to interval 1–2, and the dashed line corresponds to interval 2–3.

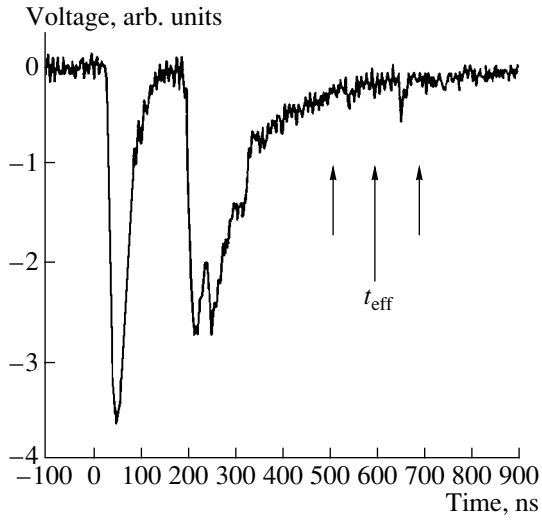


Fig. 8. Signal from neutron detector *D*2 obtained in shot no. 5.

processes (5) and (6) for a "slow" component of the liner. To verify this model of the generation of optical radiation, all three optical detectors in shots nos. 6 and 7 were shifted away from the CIS by 5 cm from their initial positions. Figure 6 shows the signals from three optical detectors in shot no. 7 and their approximations by pulse functions. Figure 7 shows the deuteron energy distribution obtained from an analysis of the signals shown in Fig. 6. It is seen that the most probable values of deuteron energies and the energy spreads obtained by data processing for interval 1–2 are in good agreement with those for interval 2–3. This coincidence of the results for two successive intervals supports the validity of the above model of the formation of optical radiation during the expansion of the deuterium liner. It is seen from Table 1 that there is a significant scatter in the liner deuteron energies (50–100%) for different shots. This may be attributed to the nonoptimal relationship between the mass of the jet in the anode–cathode gap and the current through the liner, as well as the insufficiently accurate timing of the pulses of the high-voltage switch, the pulsed-voltage generator, and the electromagnetic valve. These issues require further investigations and improvement of the method.

2. Neutron Detectors

Figure 8 shows the signal from the scintillation neutron detector *D*2 for shot no. 5. The target placed around the liner was a cylindrical sheath 36 cm in radius and 40 mm in length; its inner surface was coated with CD_2. The first spike responds to bremsstrahlung; its source may be associated with either the initial breakdown of the anode-cathode gap in the accelerator load unit (the acceleration of the deuterium-plasma electrons up to energies corresponding to the amplitude of the high-voltage pulse and their subsequent interaction with the anode material and the walls of the measuring chamber) or the formation of the magnetic insulation in the vacuum transmission line leading to the load unit. The second spike is associated with the background neutron emission arising from reaction (1) during the electrodynamic acceleration of the liner (the generation of high-energy deuterons due to the possible onset of MHD instability in the liner) [11]. The time interval between the first and second spikes characterizes the time required for a neutron from reaction (1) to fly the distance l from the liner axis to scintillation detector *D*2

$$\Delta t[\text{ns}] = t_n - t_\gamma, \tag{7}$$

where t_n and t_γ are the times it takes for a neutron and bremsstrahlung photon to fly from the liner axis to detector *D*2.

For a given value of L, the maximum energy of neutrons detected by detector *D*2 positioned at an angle of 56° to the anode–cathode direction in the load unit can be easily determined from Eq. (7), assuming that γ-ray photons and neutrons are generated simultaneously. The corresponding energy of deuterons accelerated due to MHD instabilities was calculated according to reaction (1).

The results of an analysis of the experimental data obtained with the method described above are presented in Table 2. We note that the maximum values of E_d and E_n agree with the results obtained in [16]. Note also that, according to [11], the background neutron fluxes in shots nos. 1–3, 6, and 7 (in the absence of a CD_2 target) are proportional to the ratio I_L^2/m_L (where

Table 2. Parameters of the deuteron and background neutron fluxes

Shot no.	1	2	3	4	5	6	7
t_n, ns	174 ± 3	175 ± 3	172 ± 3	181 ± 3	179 ± 3	170 ± 3	178 ± 3
E_n, MeV	2.90 ± 0.10	2.87 ± 0.10	2.97 ± 0.11	2.68 ± 0.09	2.74 ± 0.09	3.04 ± 0.11	2.77 ± 0.09
E_d, keV	+118	+113	+132	+76	+89	+141	+95
	310	274	386	103	152	470	179
	−101	−95	−112	−58	−71	−122	−77

Note: t_n is the time it takes for a neutron to fly from the liner axis to detector $D2$; E_n and E_d are the maximum neutron and deuteron energies, respectively.

I_L^2 and m_L are the current and mass of the liner, respectively). At present, the data obtained in shots nos. 4 and 5 are being analyzed in order to determine the yield of the "target" neutrons produced in the interaction of the liner deuterons with the deuterium target (the time interval corresponding to the detection of the "target" neutrons is indicated by arrows in Fig. 8).

CONCLUSIONS

The results obtained allow us to draw the following conclusions:

(i) A simple optical diagnostics for studying the liner dynamics in an inverse Z-pinch has been developed.

(ii) The characteristics of the three-channel optical detection system for measuring the ion energy distribution in the liner by the flight-of-time technique have been studied.

(iii) The maximum energies of the accelerated deuterons generated due to the onset of MHD instabilities and determining the intensity of the background neutron emission have been measured with the use of scintillation detectors.

(iv) A simultaneous analysis of the data obtained with the help of optical detectors, magnetic probes, and neutron detectors, as well as the use of the results from bolometric measurements, will make it possible to obtain precise information about the ion energy distribution in the liner.

ACKNOWLEDGMENTS

We thank our colleagues E.N. Volkov, V.I. Makhrin, and V.A. Sinebryukhov (Institute of High-Current Electronics) for the maintenance of the accelerator and assistance in measurements. This work was supported in part by the Russian Foundation for Basic Research (project no. 00-02-17203) and the Foundation of the Plenipotentiary of Poland at the Joint Institute for Nuclear Research.

REFERENCES

1. V. B. Belyaev, A. Bertin, Vit. M. Bystritskii, *et al.*, Nucleonica **40**, 85 (1995).
2. V. B. Belyaev, Vyach. M. Bystritsky, O. I. Kartavtsev, *et al.*, Preprint No. D15-92-324 (Joint Inst. for Nuclear Research, Dubna, 1992).
3. V. M. Bystritsky, V. M. Grebenyuk, S. S. Parzhitski, *et al.*, Preprint No. D15-96-11 (Joint Inst. for Nuclear Research, Dubna, 1996).
4. V. M. Bystritsky, V. M. Grebenyuk, S. S. Parzhitski, *et al.*, Nucleonica **42**, 775 (1997).
5. T. Bulgakov, Vit. M. Bystritskii, G. Mesyats, *et al.*, in *Proceedings of the 11th International Conference on High-Power Particle Beams, Prague, 1996*, Vol. 2, p. 917.
6. Vit. M. Bystritskii *et al.*, in *Albuquerque*, Ed. by W. L. Baker and G. Coperstein (IEEE, New York, 1995), p. 1215.
7. T. L. Bulgakov, Vit. M. Bystritskii, Vyach. M. Bystritsky, *et al.*, Yad. Fiz. **60**, 1345 (1997) [Phys. At. Nucl. **60**, 1217 (1997)].
8. V. M. Bystritski, V. M. Grebenyuk, S. S. Parzhitski, *et al.*, Laser Part. Beams **18**, 325 (2000).
9. V. M. Bystritskiĭ, V. M. Bystritsky, S. Chaikovsky, *et al.*, Kerntechnik **66**, 42 (2001); V. M. Bystritskii, V. M. Bystritsky, S. Chaikovsky, *et al.*, Yad. Fiz. **64**, 920 (2001) [Phys. At. Nucl. **64**, 855 (2001)].
10. V. M. Bystritsky and F. M. Pen'kov, Preprint No. D17-2000-293 (Joint Inst. for Nuclear Research, Dubna, 2000).
11. V. M. Bystritsky, V. M. Grebenyuk, S. S. Parzhitski, *et al.*, Nucl. Instrum. Methods Phys. Res. A **455**, 706 (2000).
12. A. Krauss, H. W. Becker, H. P. Trautvetter, *et al.*, Nucl. Phys. A **465**, 150 (1987).
13. N. Jarmie, R. E. Brown, R. A. Hardekopf, *et al.*, Phys. Rev. C **29**, 2031 (1984).
14. A. V. Luchinskiĭ, V. I. Makhrin, N. A. Ratakhin, *et al.*, Izv. Vyssh. Uchebn. Zaved., Fiz. **38** (12), 52 (1995).
15. C. F. Barnett and M. F. A. Harrison, in *Applied Atomic Collision Physics*, Ed. by H. S. W. Massey, E. W. McDaniel, and B. Bederson (Academic, New York, 1984).
16. R. Aiaga-Rossel *et al.*, in *Proceedings of the 4th International Conference on Dense Z-Pinches, 1997*, Ed. by N. R. Pereira, J. Davis, and P. E. Pulsifer (AIP, Woodbury, 1997), p. 61.

Translated by N. Larionova

Technical Physics, Vol. 47, No. 9, 2002, pp. 1106–1111. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 38–43.

GAS DISCHARGES, PLASMA

Statistical Study of Nanosecond Electric Breakdowns in *n*-Hexane

V. F. Klimkin

Novosibirsk State University, ul. Pirogova 2, Novosibirsk, 630090 Russia

e-mail: klimkin@phys.nsu.ru

Received February 11, 2002

Abstract—Methods for analyzing statistical distributions of the breakdown delay time are generalized. A statistical approach is used to study electric breakdown in *n*-hexane in a 2.1-MV/cm quasi-uniform electric field at a pulse duration of $\sim 5 \times 10^{-8}$ s. Two different mechanisms for the anode breakdown are shown to coexist and compete with each other. One of them incorporates the "bubble" stage, whereas the other one is related to ionization in the liquid itself. It is found that the weaker influence of the external pressure on the pulsed electric strength of liquids in the nanosecond range is caused by a transition to the ionization mechanism for the anode breakdown at elevated pressures. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The need for a detailed study of pulsed electric breakdowns in liquid dielectrics stems mainly from the wide use of these liquids in various electrophysical devices (energy storage banks, high-current switches, facilities for pulsed-discharge technologies, ionization chambers in nuclear physics, etc.). Experimental data obtained at high electric fields in the range ≥1 MV/cm at nanosecond pulse durations are very scarce. There are various hypotheses about the mechanisms for nanosecond electric breakdown in liquid, which sometimes contradict each other. For example, in [1] it was suggested that the breakdown mechanism, being electrical in nature, is induced by the ionization processes in the liquid itself. In [2], a thermal mechanism was proposed in which the primary process was assumed to be the formation of vapor-gas bubbles with their subsequent ionization. Up to now, the ionization mechanism for nanosecond breakdown is considered as arguable even in the case of the best liquid dielectrics [3]. The external pressure is one of the factors that increase the electric strength of liquid dielectrics. It was pointed out that the effect of the external pressure on the breakdown delay time in liquids decreases when passing over from microsecond to nanosecond voltage pulses [4, 5]. Interesting features of the electric breakdown in *n*-hexane in a quasi-uniform field were found in [5]. It was shown that, at the nanosecond pulse durations, two types of the anode breakdown coexist and compete with each other. The first type of breakdown starts to develop from a slow first stage (in appearance, it resembles a deformed microbubble), which then transforms into a rapidly developing brush discharge (complex electric discharge). The other type of breakdown reveals itself as a homogeneous electric discharge that does not exhibit any slow first stage. The discharge develops immediately with a high speed approximately equal to the speed of a brush discharge ($\sim(1–2) \times 10^6$ cm/s for a 100-μm-long discharge gap). By applying an external pressure, these types of breakdown can be separated in time, which indicates the different nature of the physical processes responsible for breakdown [6]. Ultrafast optical studies of a discharge developing in distilled water from a point anode confirmed the coexistence and competition between the two different mechanisms for nanosecond breakdown [7].

The statistical nature of breakdown requires the statistical treatment of experimental data. Important data on prebreakdown processes can be obtained by analyzing the probability distributions of certain characteristic parameters, such as the breakdown electric field, the breakdown delay time, and the duration of the prebreakdown glow.

This paper is devoted to the statistical study of the mechanisms for electric breakdown in *n*-hexane at voltage pulse durations of ~50–100 ns.

METHODS FOR ANALYZING STATISTICAL DISTRIBUTIONS OF THE BREAKDOWN DELAY TIME

In explaining the phenomenon of the delayed breakdown in gases, it is commonly accepted that a free electron capable of ionizing must appear near the cathode to give rise to an electron avalanche. The appearance of an electron near the cathode is a random (i.e., statistical) process. For this reason, the above phenomenon is called the statistical delay of a discharge, whereas the time lag between the instant when the voltage is applied and the appearance of a primary electron inducing the breakdown is called the statistical delay time t_s. This time differs from the discharge formation time t_f, which

1063-7842/02/4709-1106$22.00

is equal to the time interval between the appearance of a primary electron and the end of gap breakdown. Thus, the breakdown delay time in gases can be represented as the sum of t_s and t_f [8].

By analogy with the breakdowns of gas gaps, we represent the breakdown delay time in liquids as a sum of two times: the statistical delay time t_s and the discharge formation time t_f. In liquids, the former time is a time required to provide conditions required for the onset of a certain perturbation near the electrode. The formation time corresponds to the propagation time of this perturbation between the electrodes. Due to the statistical nature of the processes responsible for gap breakdown, the delay time is a fluctuating quantity.

Statistical Delay Time

The distribution of events in a series of N independent successive tests, or when simultaneously observing a set of N objects (provided that the event either occurs or not), is described by Bernoulli's binomial law [9].

Let us consider the probability of breakdown initiation over a time interval t. We will investigate the problem using the following assumptions: whether the initiation over a time interval t will occur or not does not depend on the history of the preceding events (independent events); the probability of breakdown initiation over a short time interval Δt is proportional to the interval duration (i.e., $p = \mu\Delta t$, where μ is the field-dependent probability of the event occurrence per unit time); and the probability of the simultaneous occurrence of two or more events per time interval Δt is negligibly low.

Let us divide the time interval t into the large number N of short intervals with the probability of breakdown initiation during each of them equal to p. The total probability over all N intervals is

$$P(t) = Np(1-p)^{N-1}.$$

Passing over to the limit $N \longrightarrow \infty$, we obtain a probability distribution with the density

$$f(t_s) = \mu\exp(-\mu t_s). \quad (1)$$

Thus, the statistical delay time obeys an exponential distribution [8, 10]. The mean duration of the statistical delay (mathematical expectation) is $\bar{t}_s = 1/\mu$. The dispersion of the statistical delay time is $\sigma^2(t_s) = (\bar{t}_s)^2$.

Total Delay Time

The probability that the breakdown will be initiated over a time of no less than the time t_s is

$$P(t_s, \infty) = \int_{t_s}^{\infty} f(t_s)dt_s = \exp(-t_s/\bar{t}_s). \quad (2)$$

Usually, the total delay time is measured at a fixed voltage, pressure, and interelectrode gap length. Based on the experimental results, one can determine the event rate with which breakdown occurs over a time of no less than the time t,

$$F(t, \infty) = N_t/N_0,$$

where N_t is the number of pulses at which the breakdown delay time is no less than t and N_0 is the total number of breakdowns.

It follows from Bernoulli's theorem [9] that, given a large number of tests, the event rate $F(t, \infty)$ approaches the probability of this event $F(t, \infty)$ (more exactly, converges in probability). Assuming the discharge formation time t_f to be approximately constant, expression (2) takes the form

$$N_t = N_0\exp[-(t-t_f)/\bar{t}_s]. \quad (3)$$

Hence, the dependence of $-\ln(N_t/N_0)$ on t should be linear. The slope of the line determines the mean statistical delay time $\bar{t}_s$, whereas the point where this line intersects the y axis gives the formation time t_f. This is a base for the Laue method for analyzing the distribution of the breakdown delay times in gases [8].

Under actual experimental conditions, the formation time t_f fluctuates. The quantity Δt_f related to the scatter in the formation times can be represented as a sum of a sufficiently large number of independent (or weakly dependent) elementary deviations (caused by different factors), each of which relatively slightly affects the sum. According to the central limit theorem, the distribution law for the sum of a large number of independent random variables, whatever distribution law they obey (provided that the above soft constraints are met), approximately corresponds to the normal distribution [9].

If the mean statistical delay time satisfies the inequality $\bar{t}_s \ll \bar{t}_f$, then the distribution of the formation time $\bar{t}_f$ can be determined experimentally. The measurements of the breakdown delay times in *n*-hexane at high electric fields (in the nanosecond pulse-duration range) showed that the distribution of the formation time is indeed close to the normal distribution [11].

The normal distribution is characterized by the probability density

$$f(t_f) = (1/\sigma\sqrt{2\pi})\exp[-(t_f-\bar{t}_f)^2/2\sigma^2], \quad (4)$$

where $\bar{t}_f$ and σ^2 are the mean value (mathematical expectation) and the dispersion of t_f, respectively.

In this case, the distribution of the total breakdown delay time can be described by a mathematical model based on the superposition of the exponential distribution of the statistical delay time and the normal distribution of the formation time [12]. Let us find the distribution of the breakdown delay time caused by two con-

secutive processes: the onset and propagation of the perturbation. We suppose that these processes are statistically independent. Let the former process (with the mean statistical delay time $\bar{t}_s$) occur within the interval dt_s. When this process is finished, the second process with the mean formation time $\bar{t}_f$ develops. We suppose that the breakdown occurs within the interval dt. The probability that the breakdown will occur within the time interval $(t, t + dt)$ is determined by the product of the probabilities for these two processes

$$dP(t, t+dt) = (1/\bar{t}_s)\exp[-t_s/\bar{t}_s]dt_s$$
$$\times (1/\sigma\sqrt{2\pi})\exp[-(t-t_s-\bar{t}_f)^2/2\sigma^2]dt.$$

For the probability of breakdown over the time interval $(0, t)$, we have

$$P(0,t) = (1/\bar{t}_s)\int_0^t \exp(-t_s/\bar{t}_s) \times \left\{(1/\sigma\sqrt{2\pi})\int_{t_s}^t \exp[-(t-t_s-\bar{t}_f)/2\sigma^2]dt\right\}dt_s. \quad (5)$$

Integrating expression (5) numerically and going over to the opposite event, we find the probability $t(P(t, \infty))$ that gap breakdown will occur over a time of no less than t.

As an example, Fig. 1 shows the computation results in the Laue coordinates (on a semilogarithmic scale) for the following distribution parameters: $\bar{t}_s$ = 100 ns, $\bar{t}_f$ = 150 ns, and σ = 25 ns. The left straight line shows the distribution of the statistical delay time. It is seen that the total distribution (the right curve) does not obey the linear law. However, the mean statistical delay time $\bar{t}_s$ and the mean formation time $\bar{t}_f$ can be determined from the linear segment of this curve. The computation

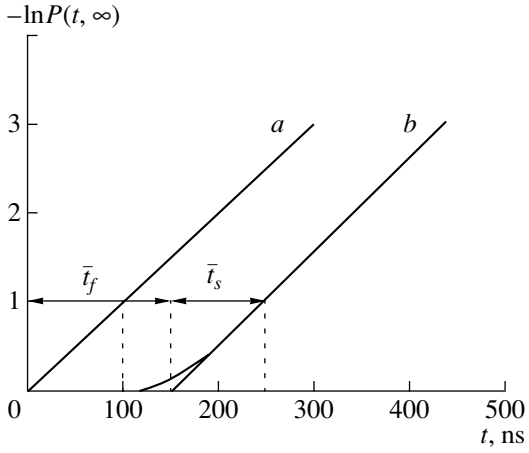


Fig. 1. Distributions of the (*a*) statistical and (*b*) total delay times of a breakdown caused by two consecutive processes.

results corresponding to the other distribution parameters show that, if $\bar{t}_s \gg \bar{t}_f$, then the total distribution differs slightly from the exponential one; otherwise, it is close to the normal distribution.

Parallel Processes

A method for analyzing statistical distributions of the breakdown delay time in the case of parallel processes was proposed in [13]. Due to the importance of the results obtained and absence of readily available literature on this problem, we will describe this method in more detail.

We consider two competing, statistically independent events, e.g., gap breakdowns starting from the anode or the cathode (or two breakdowns of different types starting from the same electrode). For the sake of simplicity, we suppose that the formation times for these processes are slightly fluctuating quantities. The probability that the first event will occur over a time of no less than t is determined by the expression

$$P_1(t, \infty) = \exp[-(t-t_{f1})/\bar{t}_{s1}].$$

Similarly, for the second event we have

$$P_2(t', \infty) = \exp[-(t'-t_{f2})/\bar{t}_{s2}].$$

Let $t_{f2} > t_{f1}$. We note that

$$P_1(t) = 0 \text{ for } t \le t_{f1}, \text{ and } t' = \begin{cases} t_{f2} & \text{for } t \le t_{f2} \\ t & \text{for } t > t_{f2}. \end{cases}$$

The probability of the gap breakdown occurring over a time interval $\ge t$ is the probability of the occurrence of the first and the second events. The total probability of the occurrence of two statistically independent events is equal to the product of the probabilities of these events (probability product rule [9])

$$P(t, \infty) = \exp[-(t-t_{f1})/\bar{t}_{s1}]\exp[-(t'-t_{f2})/\bar{t}_{s2}]. \quad (6)$$

Taking the logarithm of expression (6), we obtain

$$-\ln P(t, \infty) = (t-t_{f1})/\bar{t}_{s1} + (t'-t_{f2})/\bar{t}_{s2}. \quad (7)$$

As an example, Fig. 2 (curve *c*) presents the calculated total probability distribution in the Laue coordinates for the following parameters: t_{f1} = 100 ns, $\bar{t}_{s1}$ = 100 ns, t_{f2} = 150 ns, and $\bar{t}_{s2}$ = 60 ns. Here, the distributions of the individual probabilities of the first (curve *a*) and second (curve *b*) events are also shown. One can see a kink on the total distribution (curve *c*) at $t = t_{f2}$. The linear segment of the curve corresponding to the time interval (t_{f1}, t_{f2}) characterizes the first event, whereas at $t > t_{f2}$, the curve describes the simultaneous occurrence of the two events. From the slope of the former segment, one can determine the mean statistical delay time $\bar{t}_{s1}$. From the slope of the latter segment, it is possible to determine the mean statistical delay time $\bar{t}_{s2}$.

Indeed, at $t > t_{f2}$, expression (7) takes the form

$$-\ln P(t, \infty) = [(\dot{t}_{s1} + \dot{t}_{s2})/\dot{t}_{s1}\dot{t}_{s2}]t - (t_{f1}\dot{t}_{s2} + t_{f2}\dot{t}_{s1})/\dot{t}_{s1}\dot{t}_{s2},$$

where

$$1/\tau = (\dot{t}_{s1} + \dot{t}_{s2})/\dot{t}_{s1}\dot{t}_{s2} = 1/\dot{t}_{s1} + 1/\dot{t}_{s2} \quad (8)$$

is the slope of the segment describing the simultaneous occurrence of the two events.

From (8), we obtain

$$\dot{t}_{s2} = \tau\dot{t}_{s1}/(\dot{t}_{s1} - \tau). \quad (9)$$

Thus, by analyzing the total distribution of the breakdown delay time, one can estimate the parameters of the processes responsible for the different types of breakdown.

EXPERIMENTAL SETUP

The experimental setup consisted of a discharge chamber and pulsed voltage generator. A manual hydraulic press of a piston type was used to produce elevated pressures in the discharge chamber. The distance between the two hemispherical 5-mm-radius stainless-steel electrodes of the discharge gap was ~100 μm. A movable grounded electrode system allowed us to smoothly regulate the gap length at a constant pressure in the chamber. An automatic optical measurement system was employed to monitor the gap length by using the Fraunhofer diffraction patterns. The relative error in measuring the gap length was ~5%. The amplitude of the voltage pulse applied to the electrodes was ~10–30 kV, the rise time was ~10 ns, and the duration of the flat-top part was ~5 μs. A processing system based on an Elektronika-60 microcomputer with a KAMAK interface [14] was used to obtain the statistical distributions of the breakdown delay time in the course of the experiment. The time measurement error was 10 ns. Simultaneously with an analysis of statistical distributions, prebreakdown phenomena in the discharge gap were visualized with the help of a laser schlieren system. A ruby laser with a pulse duration of ~5 ns was used as an illumination source. Experiments were carried out with chemically pure *n*-hexane (with a specific resistance of $\geqslant 10^{12}$ Ω cm); no outgassing of the electrodes and liquid was performed.

EXPERIMENTAL RESULTS AND DISCUSSION

Figure 3 presents typical histograms of the distribution of the breakdown delay time in *n*-hexane at an electric field strength of ~2.1 MV/cm (which corresponds to a nanosecond breakdown) and different external pressures. It is seen that the increase in the pressure changes the distribution shape from a nearly symmetric (at atmospheric pressure) to fairly asymmetric (at P = 0.9 MPa) profile. The mean breakdown delay time $\dot{t}$ also increases with pressure.

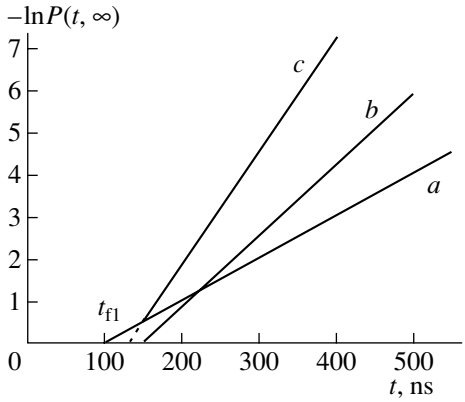


Fig. 2. (*a*, *b*) Probability distributions for two independent events and (*c*) the total probability distribution for two parallel processes.

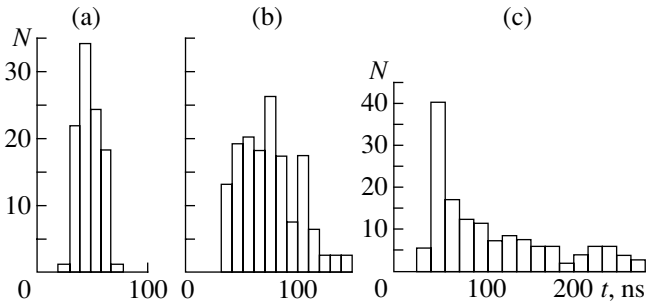


Fig. 3. Distributions of the breakdown delay time for (a) $\dot{t}$ = 53 ns, σ = 11 ns, P = 1 atm, and N_0 = 100; (b) $\dot{t}$ = 79 ns, σ = 28 ns, P = 0.5 MPa, and N_0 = 150; and (c) $\dot{t}$ = 100 ns, σ = 68 ns, P = 0.9 MPa, and N_0 = 136. Here, σ is the root-mean-square deviation of an individual measurement and N_0 is the total number of measurements; the interelectrode distance is 100 μm.

The distributions of breakdown delay time obtained by the Laue method for different external pressures are shown in Fig. 4. A characteristic feature of the distributions is the appearance of a kink at elevated pressures (Fig. 4, curves *3*, *4*). This behavior of the Laue curves can be described with the help of a mathematical model based on the superposition of two exponential distributions corresponding to the competing, statistically independent events with different characteristic times. Thus, a kink on the Laue curve points to the possible coexistence of two independent competing breakdown mechanisms. A comparison with the results of optical monitoring shows that breakdown occurs via the above-discussed electric discharges starting from the anode. A subsequent smooth transition to a straight line (Fig. 4, curve *5*) indicates that the increase in the external pressure changes the breakdown mechanism.

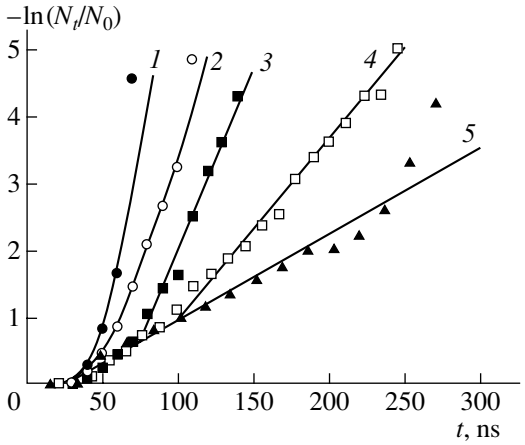


Fig. 4. Distributions of the breakdown delay time in the Laue coordinates at an electric field of 2.1 MV/cm for (*1*) P = 1 atm, $\dot{t}$ = 53 ns, and N_0 = 100; (*2*) P = 0.3 MPa, $\dot{t}$ = 63 ns, and N_0 = 130; (*3*) P = 0.5 MPa, $\dot{t}$ = 79 ns, and N_0 = 150; (*4*) P = 0.7 MPa, $\dot{t}$ = 88 ns, and N_0 = 150; and (*5*) P = 0.9 MPa, $\dot{t}$ = 100 ns, and N_0 = 136. N_0 is the total number of measurements.

By analyzing the total distributions of the breakdown delay time (for successive and parallel processes), one can estimate the mean statistical delay time $\dot{t}_s$ and the mean formation time $\dot{t}_f$ for each of the breakdown mechanisms and determine how they depend on the external pressure. These data are presented in Fig. 5. The strong influence of the pressure on the first breakdown mechanism points to the important role of the bubble stage during the bubble origin and formation. At the same time, the fact that the pressure does not affect the second breakdown mechanism indicates that the bubble stage does not occur (or is not dominant) in this case. The average rate of the electric discharge development increases (due to the absence of the bubble stage) and amounts to ~4×10^5 cm/s.

Figure 6 shows the mean breakdown delay time deduced from analyzing statistical distributions ($\dot{t} = \dot{t}_s + \dot{t}_f$) for the first (▲) and second (□) mechanisms as functions of the external pressure. Here, the results of direct measurements with the help of the automatic system are also shown (●). Every point is an average of 100–150 measurements. It is seen that, at atmospheric (or slightly higher) pressure, the first mechanism (corresponding to a complex discharge) mainly occurs. In this case, $\dot{t}_f > \dot{t}_s$ and the breakdown delay time has a nearly normal distribution (Fig. 3a). An increase in the pressure increases the breakdown delay time corresponding to the first mechanism, which leads to the coexistence of and competition between the two different breakdown mechanisms. At elevated pressures of $P \geq 0.9$ MPa, the second breakdown mechanism (it corresponds to a homogeneous discharge) mostly occurs. In this case, $\dot{t}_s > \dot{t}_f$ and the breakdown delay time has a nearly exponential distribution (Fig. 3c). Thus, the weaker influence of the external pressure on the mean breakdown delay time is related to the transition to the ionization mechanism for the anode breakdown (homogeneous discharge) at elevated pressures.

The vertical dashed line in Fig. 5 shows the pressure at which a complex electric discharge does not have time to form, and a transition to a homogeneous electric discharge occurs. The strong influence of a relatively low external pressure on the statistical delay time of the complex electric discharge starting from the anode indicates the thermal mechanism for the origin of the bubble stage. Hence, the statistical delay time of the complex electric discharge incorporates a process that leads to the rapid local heating of the liquid near the anode surface and the formation of a vapor microbub-

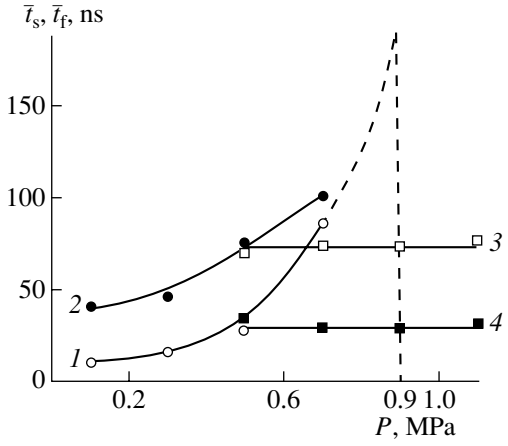


Fig. 5. Mean statistical delay time and mean formation time vs. external pressure for the breakdowns of the first (*1*, *2*) and second (*3*, *4*) type.

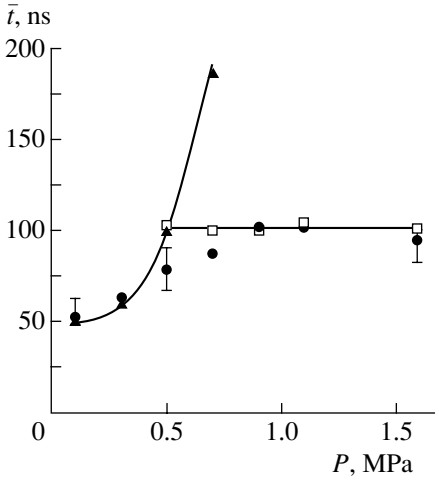


Fig. 6. Mean breakdown delay time vs. external pressure for the first (▲) and second (□) breakdown mechanisms; the results from direct measurements of the breakdown delay time are also shown (●).

ble. The formation time is determined by the bubble growth, the development of ionization processes inside it, and the initiation of the fast ionization processes in the liquid itself due to the increased electric field near the pole of the deformed bubble (which becomes conducting after breakdown). For a homogeneous electric discharge starting from the anode, the statistical delay time is the time it takes for the self-sustaining ionization processes in the liquid to begin. Under these conditions, not only the presence of sharp micropoints on the anode surface, but also the properties of the liquid near the point ends become especially important. The formation time corresponds to the propagation of the ionization processes from the anode to cathode.

CONCLUSION

The results obtained allow us to close the discussion between the supporters of the bubble and electric mechanisms for nanosecond breakdowns. It turns out that both of these mechanisms can operate. Which one prevails is determined by experimental parameters such as the electric field, the interelectrode distance, and the external pressure.

REFERENCES

1. V. Ya. Ushakov, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 1, 105 (1979).
2. É. V. Yanshin, I. T. Ovchinnikov, and Yu. N. Vershinin, Dokl. Akad. Nauk SSSR **214**, 1303 (1974).
3. S. M. Korobeĭnikov, Teplofiz. Vys. Temp. **36**, 541 (1998).
4. K. Arii and I. Kitani, J. Phys. D **14**, 1675 (1981).
5. V. P. Borodin and V. F. Klimkin, Pis'ma Zh. Tekh. Fiz. **14**, 802 (1988) [Sov. Tech. Phys. Lett. **14**, 358 (1988)].
6. V. F. Klimkin, Zh. Tekh. Fiz. **60** (6), 161 (1990) [Sov. Phys. Tech. Phys. **35**, 735 (1990)].
7. V. F. Klimkin, Pis'ma Zh. Tekh. Fiz. **16** (4), 54 (1990) [Sov. Tech. Phys. Lett. **16**, 146 (1990)].
8. J. M. Meek and J. D. Craggs, *Electrical Breakdown of Gases* (Clarendon, Oxford, 1953; Inostrannaya Literatura, Moscow, 1965).
9. E. S. Venttsel', *Probability Theory* (Nauka, Moscow, 1969).
10. T. J. Lewis and B. W. Ward, Proc. R. Soc. London, Ser. A **269**, 233 (1962).
11. A. J. Beddow and J. E. Brignell, Electron. Lett. **2** (4), 142 (1966).
12. A. J. Beddow and J. E. Brignell, Electron. Lett. **1** (9), 253 (1965).
13. A. P. Avrorov and V. V. Vorob'ev, Preprint No. 83-69 (Institute of Nuclear Physics, Siberian Division, USSR Academy of Sciences, Novosibirsk, 1983).
14. V. F. Klimkin, Zh. Tekh. Fiz. **62** (9), 160 (1992) [Sov. Phys. Tech. Phys. **37**, 965 (1992)].

Translated by N. Ustinovskiĭ

Technical Physics, Vol. 47, No. 9, 2002, pp. 1112–1119. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 44–51.

SOLIDS

Skin Effect in Massive Conductors Used in Pulsed Electrical Devices: I. Electromagnetic Field of Massive Conductors

B. E. Fridman

Institute of Problems in Electrophysics, Russian Academy of Sciences, St. Petersburg, 191186 Russia
e-mail: fridman@mail.infostar.ru

Received June 27, 2001; in final form, February 21, 2002

Abstract—The diffusion of a pulsed electromagnetic field into massive conductors with an arbitrary smooth surface is considered for the case where the field penetration depth is small. By using the boundary layer method, an asymptotic solution for the electromagnetic field is constructed. First- and second-order corrections to the limiting solution, which corresponds to the field distribution at an indefinitely high conductivity of the conductors, are found. Time dependences of the first- and second-order approximations to the electric field on the surface of the conductor are determined. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Massive conductors withstanding currents as high as several megaamperes or several tens of megaamperes are a basic component of devices generating high current pulses [1]. The depth of penetration of a pulsed magnetic field into these conductors is much smaller than their characteristic sizes; that is, their skin depth is small.

The skin effect in massive conducting bodies has been the subject of much investigation. When the skin depth is small, an external quasi-stationary variable electromagnetic field is usually calculated with approximate techniques, using either of two approximate boundary conditions. In the former case, the conductivity of the conductor is assumed to be indefinitely high and the tangential component of the electric field vector and the normal component of the magnetic field vector on the surface are zero (perfect conductivity conditions). In the latter case, impedance boundary conditions are used. It is assumed here that a local electromagnetic wave penetrates into a metal in the same way that a plane wave penetrates into a conductive half-space [2]. For nonferromagnetic metals, the external field turns out to differ insignificantly from that obtained with the boundary conditions of the first type. This provides a way of analyzing systems with massive conductors by means of two successive approximations [3]. In the first approximation, one finds the distribution of an external magnetic field and the linear current density in the conductor conditions of perfect conductivity. In the second approximation, it is assumed that the local distributions of the field and current at the surface correspond to the penetration of a 1D field into a conductive half-space (that is, the impedance boundary conditions are used).

One can evaluate the field numerically by directly using the impedance boundary conditions. In this case, the boundary element method [4, 5] is employed and the field in the outer region is described by integral equations including the boundary conditions.

A relatively general approach to calculating variable and pulsed electromagnetic fields in conductive bodies was used in [6–9]. In [6], the field is determined in a local system of curvilinear coordinates that is placed on the surface of a conductive sheath of arbitrary geometry. Such calculations are aimed at finding the field penetrating into the sheath. Therefore, the field excited by external current sources connected to a conductive body was not considered.

To calculate the field when the skin depth is very small, it is natural to take advantage of the boundary layer method. Such an approach to determining the field and to calculating the heating of a conductor by eddy currents dates back to 1940 [7]. The boundary layer method was also applied to finding boundary conditions on the conductor surface. The conditions found were then used in integral equations of the boundary element method [8, 9].

However, the problem of deriving general relationships for the current-excited field in massive conductors still remains urgent. This problem can be resolved for the case of the sharp skin effect, when an electromagnetic field propagates in a thin near-surface layer. Such a situation is typical of sources and loads of high current pulses. In this case, one can specify local curvilinear coordinates in the near-surface layer of a metal and find asymptotic dependences between the electric and magnetic field vectors on the surface of conductors of arbitrary geometry. The purpose of this work is to derive and substantiate such dependences.

1063-7842/02/4709-1112$22.00

BASIC RELATIONSHIPS

Let us consider a set of massive metallic conductors with a current pulse $I(t)$. The electromagnetic field in the conductors (region Ω_i) and in the environment (region Ω_e) is described by the equation for magnetic induction vector $\mathbf{B}$:

$$\Delta\mathbf{B} = \alpha\mu_0\gamma\frac{\partial\mathbf{B}}{\partial t}, \quad (1)$$

where $\alpha = 0$ and 1 outside and inside the metal, respectively; γ is the conductivity of the metal; t is time; $\mu_0 = 4\pi \times 10^{-7}$ H/m; and Δ is Laplacian.

If the skin depth is small, the magnetic induction vector can be divided into two components:

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_S, \quad (2)$$

where $\mathbf{B}_0$ is the electromagnetic field for the perfect skin effect ($\gamma = \infty$) and $\mathbf{B}_S$ takes into account the effect of eddy currents in the metal on the resultant field.

At points lying on the surface Γ of the conductors, the vector $\mathbf{B}_0$ is directed tangentially to the metal surface:

$$[\mathbf{n} \times \mathbf{B}_0]_\Gamma = \mu_0\boldsymbol{\delta}, \quad (3)$$

where $\boldsymbol{\delta}$ is the linear current density in the conductor in the limiting case (the perfect skin effect) and $\mathbf{n}$ is the outer normal to the surface Γ.

The field $\mathbf{B}_0$ defines the external inductance L of the set of conductors. The calculation of the electromagnetic field for the perfect skin effect and the determination of the inductance L have been the subject of a variety of works (see, e.g., [3, 10]). In this work, we do not consider related computing methods. The property of the field $\mathbf{B}_0$ we are interested in is that time t and space variables can be separated in this field. Then, the linear current density δ can be represented as

$$\boldsymbol{\delta}(\mathbf{r}, t) = \mathbf{b}(\mathbf{r})I(t), \quad (4)$$

where $\mathbf{r}$ is the radius vector of a point on the surface Γ, $\mathbf{b}(\mathbf{r})$ is a vector function that is specified on the surface Γ and depends only on the conductor geometry, and $I(t)$ is the current in the conductors.

ELECTROMAGNETIC FIELD IN THE SURFACE LAYER OF THE CONDUCTOR

We consider conductors with a smooth surface; that is, the minimal radius of curvature is much larger than the penetrating depth of the electromagnetic field. In the surface layer of the conductor, where the field is calculated, the inner normals to the surface do not intersect. Let us introduce coordinates (x_1, x_2) along the lines of curvature on the surface Γ. Such coordinates are orthogonal, and the first and second quadratic forms of the surface Γ are expressed as [11]

$$d\mathbf{r}^2 = h_{10}^2 dx_1^2 + h_{20}^2 dx_2^2,$$

$$-d\mathbf{r} \cdot d\mathbf{n} = k_1 h_{10}^2 dx_1^2 + k_2 h_{20}^2 dx_2^2,$$

where $\mathbf{r}(x_1, x_2)$ is the radius vector of a point (x_1, x_2) on the surface Γ, and k_1 and k_2 are the principal curvatures of the surface Γ at the point (x_1, x_2).

Let us introduce the third spatial coordinate x_3 into a layer adjacent to the surface Γ in such a way that the x_3 axis is aligned with the inner normal to the metal surface. Thus, we are dealing with the orthogonal coordinate system (x_1, x_2, x_3) in the vicinity of surface points for which the Lamé coefficients are given by

$$h_1 = h_{10}(1 - k_1 h_3 x_3) = h_{10}(1 - \varepsilon\chi_1 x_3),$$

$$h_2 = h_{20}(1 - k_2 h_3 x_3) = h_{20}(1 - \varepsilon\chi_2 x_3),$$

$$h_3 = \sqrt{\frac{t_1}{\mu_0\gamma}}.$$

Here, t_1 is the characteristic time of a pulsed process (e.g., the current pulse duration); $\varepsilon = h_3 k \ll 1$ is a small parameter; $K = \max_\Gamma(|k_1|, |k_2|)$ or $K = 1/a$, where a is the characteristic size of the conductor, which does not exceed the radius of curvature of the surface Γ; $\chi_1 = k_1/K$; and $\chi_2 = k_2/K$. The coordinate system (x_1, x_2, x_3) on the surface Γ is assumed to be right-handed.

Hereafter, the superscript Γ will refer to field vectors and their components at points on the metal surface. The first subscript i will designate the related coordinate from the local system. It equals 1 or 2, unless otherwise stated. The second subscript, if any, at field components indicates the serial number of the components of asymptotic expansions.

It is assumed that all the functions considered (e.g., the components of the vector $\mathbf{B}$ and their derivatives) are sufficiently smooth. This implies that any function $X(x_1, x_2, x_3)$ changes insignificantly when its arguments x_1 and x_2 are changed by h_3:

$$\left|\frac{h_3}{h_i}\frac{\partial X}{\partial x_i}\right| < \varepsilon|X|.$$

Then, the derivatives of the Lamé coefficients h_1 and h_2 and the function $X(x_1, x_2, x_3)$ are estimated as

$$\left|\frac{\partial h_i}{\partial x_{i'}}\right| \le K h_{i0} h_{i'0}, \quad \left|\frac{\partial X}{\partial x_i}\right| \le h_{i0} K|X|,$$
$$\left|h_i\frac{\partial X}{\partial x_{i'}}\right| \le 2h_{i0}h_{i'0}K|X|, \quad (5)$$

where $i, i' = 1, 2$.

For the region Ω_i of the conductor, expressions (1) and (2) and inequalities (5) yield the following relationships for the Laplacian components $(\Delta\mathbf{B}_S)_i$ in the local

coordinate system:

$$h_3^2(\Delta \mathbf{B}_S)_i = \frac{\partial^2 B_i}{\partial x_3^2}$$

$$-\varepsilon\left(\frac{\chi_1}{1-\varepsilon\chi_1 x_3}+\frac{\chi_2}{1-\varepsilon\chi_2 x_3}\right)\frac{\partial B_i}{\partial x_3}+\varepsilon\Phi_i = \frac{\partial B_i}{\partial \tau}, \quad (6)$$

$$h_3^2(\Delta \mathbf{B}_S)_3 = \varepsilon\Phi_3 = \frac{\partial B_3}{\partial \tau},$$

where B_1, B_2, and B_3 are the components of the vector $\mathbf{B}_S$ and $\tau = t/t_1$. For the functions Φ_1, Φ_2, and Φ_3, we have

$$|\Phi_i| \le \left|\frac{\partial B_3}{\partial x_3}\right| + \varepsilon[|B_i|(2+|\chi_1\chi_2-\chi_i^2|)+2|B_{3-i}|+2|B_3|],$$

$$|\Phi_3| \le 2\left(\left|\frac{\partial B_1}{\partial x_3}\right|+\left|\frac{\partial B_2}{\partial x_3}\right|\right)+\varepsilon(2|B_1|+2|B_2|+2|B_3|).$$

The electric field tangential components on the conductor surface $\mathbf{E}^\Gamma$ in the local coordinate system are found from Ampere's circuilal law $\mathbf{E} = (\mu_0\gamma)^{-1}\nabla\times\mathbf{B} = h_3^2 t_1^{-1}\nabla\times\mathbf{B}$:

$$E_{3-i}^\Gamma = (-1)^{3-i}\frac{h_3}{t_1}\left[\frac{\partial B_i}{\partial x_3}-\varepsilon\chi_i B_i-\frac{h_3}{h_{i0}}\frac{\partial B_3}{\partial x_i}\right]_\Gamma. \quad (7)$$

Following asymptotic methods for solving boundary-layer problems [12], we represent the electromagnetic field components as series in powers of the small parameter ε:

$$B_i = B_{i,0}+\varepsilon B_{i,1}+\varepsilon^2 B_{i,2}+\ldots, \quad (8)$$

$$E_i = E_{i,0}+\varepsilon E_{i,1}+\varepsilon^2 E_{i,2}+\ldots, \quad (9)$$

where $i = 1, 2, 3$.

Substituting series (8) into (6) and equating terms with equal powers of ε, we come to the set of 1D boundary-value problems

$$\frac{\partial^2 B_{i,0}}{\partial x_3^2}-\frac{\partial B_{i,0}}{\partial \tau} = 0, \quad B_{i,0}\big|_{x_3=0} = (-1)^i\mu_0\delta_{3-i}, \quad (10)$$

$$B_{i,0}\big|_{x_3\to 0} \longrightarrow 0, \quad B_{i,0}\big|_{\tau=0} = 0,$$

$$\frac{\partial B_{i,1}}{\partial x_3^2}-\frac{\partial B_{i,1}}{\partial \tau} = (\chi_1+\chi_2)\frac{\partial B_{i,0}}{\partial x_3}, \quad (11)$$

$$B_{i,1}\big|_{x_3=0} = B_{i,1}^\Gamma, \quad B_{i,1}\big|_{x_3\to\infty} \longrightarrow 0, \quad B_{i,1}\big|_{\tau=0} = 0.$$

Conditions for $B_{i,0}\big|_{x_3=0}$ are defined in (10) in view of (3). Differential equations (11) take into account that the normal component of the magnetic induction $B_{3,0} = 0$ in the first approximation.

Substituting (9) into (7) and equating terms with equal powers of ε, we get

$$E_{i,0}^\Gamma = (-1)^i\frac{h_3}{t_1}\frac{\partial B_{3-i,0}}{\partial x_3}\bigg|_\Gamma, \quad (12)$$

$$E_{i,1}^\Gamma = (-1)^i\frac{h_3}{t_1}\left(\frac{\partial B_{3-i,1}}{\partial x_3}\bigg|_\Gamma-\chi_{3-i}B_{3-i,0}^\Gamma\right). \quad (13)$$

DIFFUSION OF ELECTROMAGNETIC FIELD INTO A METAL

The application of the integral Fourier transformation in time τ to (10) yields two boundary-value problems for ordinary differential equations. A solution to them gives the Fourier transform for two tangential components of the magnetic induction in the first approximation:

$$\hat{B}_{i,0} = (-1)^i\mu_0\hat{\delta}_{3-i}\exp(-\beta x_3), \quad (14)$$

$$\beta = \begin{cases}\sqrt{\dfrac{\omega}{2}}(1+j) & \text{for} \quad \omega\ge 0,\\ \sqrt{-\dfrac{\omega}{2}}(1-j) & \text{for} \quad \omega\le 0,\end{cases}$$

where $\hat{B}_{i,0}$ and $\hat{\delta}_i$ are the Fourier transforms of the functions $B_{i,0}$ and δ_i, respectively, and $j = \sqrt{-1}$.

In boundary-value problems (11), the boundary conditions $B_{i,1}\big|_{x_3=0} = B_{i,1}^\Gamma$ must be found from joining conditions at the boundary Γ, i.e., from the condition that the components of the vector $\mathbf{B}$ in Ω_i and Ω_e equal each other at the boundary. In the second approximation (for the coefficients at ε^1 in asymptotic expansions (8) and (9)), the external field (in Ω_e) is generated by the normal component $B_{3,1}^\Gamma$ of the induction and depends on its value at the boundary Γ. The boundary value of the normal component $B_{3,1}^\Gamma$, in its turn, can be found by expanding (in powers of ε) the continuity equation for magnetic induction flux, $\nabla\cdot\mathbf{B} = 0$, in the metal region Ω_i:

$$\frac{\partial B_{3,1}}{\partial x_3} = -\frac{1}{Kh_{10}h_{20}}\left[\frac{\partial(h_{20}B_{1,0})}{\partial x_1}+\frac{\partial(h_{10}B_{2,0})}{\partial x_2}\right].$$

Hence, in view of (14), the Fourier transform of the magnetic induction normal component at the boundary is

$$\hat{B}_{3,1}^\Gamma = \frac{\mu_0}{\beta Kh_{10}h_{20}}\left[\frac{\partial(h_{10}\hat{\delta}_1)}{\partial x_2}-\frac{\partial(h_{20}\hat{\delta}_2)}{\partial x_1}\right]. \quad (15)$$

Now, let us find in Ω_e the scalar potential U_1 for the vector field $\mathbf{B}_1$, which is correct to the second order: $\nabla U_1 = -\mathbf{B}_1$. The potential U_1 can be found as a solution of the external Neumann problem for the Laplace equation in Ω_e:

$$\Delta U_1 = 0, \quad \left.\frac{\partial U_1}{\partial n}\right|_\Gamma = -B^\Gamma_{3,1}.$$

From the values of U_1, one can find the tangential components of the second-order correction to the magnetic induction:

$$B^\Gamma_{i,1} = -\frac{1}{h_{i0}}\left.\frac{\partial U_1}{\partial x_i}\right|_\Gamma. \tag{16}$$

It should be noted that the frequency (ω) dependences of the functions $\beta\hat{B}^\Gamma_{3,1}$ and $\beta\hat{B}^\Gamma_{i,1}$ are defined by the factors $\hat{\delta}_i(\omega) = b_i(\mathbf{r})\hat{I}(\omega)$. Therefore, we can write

$$\beta\hat{B}^\Gamma_{i,1} = \mu_0 K^{-1} f_i(\mathbf{r})\hat{I}(\omega), \tag{17}$$

where $f_i(\mathbf{r})$ is a function dependent only on the position of a point on the surface Γ and $\hat{I}(\omega)$ is the Fourier transform of the total current $I(\tau)$.

Applying the Fourier transformation in time τ to (11), we reduce the second-order calculations to the solution of ordinary differential equations

$$\frac{d^2\hat{B}_{i,1}}{dx_3^2} - \beta^2\hat{B}_{i,1} = (\chi_1 + \chi_2)\frac{d\hat{B}_{i,0}}{dx_3},$$

$$\hat{B}_{i,1}\big|_{x_3=0} = \hat{B}^\Gamma_{i,1}, \quad \hat{B}_{i,1}\big|_{x_3\to\infty} \longrightarrow 0.$$

Their solutions are given by

$$\begin{aligned}\hat{B}_{i,1} &= \hat{B}^\Gamma_{i,1}\exp(-\beta x_3)\\ &+ (-1)^i\frac{\chi_1+\chi_2}{2}\mu_0\hat{\delta}_{3-i}x_3\exp(-\beta x_3).\end{aligned} \tag{18}$$

Solution (18) shows that in the second approximation the magnetic induction inside the metal depends on the external field nonuniformity (the first term in (18)) and the curvature of the metal surface (the second term in (18)).

With (12)–(14) and (18), we find the Fourier transforms of the electric field components on the surface Γ in the first and second approximations:

$$\hat{E}^\Gamma_{i,0} = \sqrt{\frac{\mu_0}{\gamma t_1}}\beta\hat{\delta}_i = \sqrt{\frac{\mu_0}{\gamma t_1}}\beta b_i\hat{I}(\omega), \tag{19}$$

$$\begin{aligned}\hat{E}^\Gamma_{i,1} &= (-1)^i\sqrt{\frac{\mu_0}{\gamma t_1}}\left(\frac{\chi_1-\chi_2}{2}\hat{\delta}_i - \frac{1}{\mu_0}\beta\hat{B}^\Gamma_{3-i,1}\right)\\ &= (-1)^i\sqrt{\frac{\mu_0}{\gamma t_1}}\left(\frac{\chi_1-\chi_2}{2}b_i - K^{-1}f_{3-i}\right)\hat{I}(\omega).\end{aligned} \tag{20}$$

Note that relationship (19) in view of (3) can be written in vector form:

$$\mathbf{E}^\Gamma_0 = \sqrt{\frac{\mu_0}{\gamma t_1}}\beta[\mathbf{n}\times\hat{\mathbf{B}}_0]_\Gamma.$$

At $\omega \geq 0$, we obtain impedance boundary conditions.

A second-order approximation for $E^\Gamma_{i,1}$ close to (20) was derived by Rytov [7]. In [7], the solution was constructed in arbitrary local orthogonal coordinates on the conductor surface that were not related to the lines of curvature. It should be emphasized that coordinates on the surface Γ must be congruent with the lines of curvature; otherwise, the orthogonality of the coordinates breaks inside the metal. Rytov's formula [7] (similar to (20)) contains the derivatives of metric coefficients with respect to surface coordinates instead of the relative curvatures χ_1 and χ_2 in (20).

In [8, 9], boundary conditions were determined in higher order (ε^2 and ε^3) approximations for calculating the field with the boundary-element method and relationships were set between the electric and magnetic vectors in the second approximation. However, the term allowing for a disturbance of the external field in Ω_e (the second term in parentheses in (20)) is absent in those formulas. The inclusion of this component in the relationships between the magnetic and electric field vectors gives no way of applying the boundary element method for calculating the field of eddy currents, as was done in [8, 9].

Applying the inverse Fourier transformation to first approximations (19), we obtain

$$\begin{aligned}E^\Gamma_{i,0} &= \sqrt{\frac{\mu_0}{\pi\gamma t_1}}\frac{\partial}{\partial\tau}\int_0^\tau\frac{\delta_i(\Theta)}{\sqrt{\tau-\Theta}}d\Theta\\ &= \sqrt{\frac{\mu_0}{\pi\gamma t_1}}b_i\frac{\partial}{\partial\tau}\int_0^\tau\frac{I(\Theta)}{\sqrt{(\tau-\Theta)}}d\Theta.\end{aligned} \tag{21}$$

Since the coefficient at $\hat{I}(\omega)$ in (20) is ω-independent, the inverse Fourier transform for the second approximations has the form

$$\begin{aligned}E^\Gamma_{i,1} &= (-1)^i\sqrt{\frac{\mu_0}{\gamma t_1}}\left[\frac{\chi_1-\chi_2}{2}\delta_i(\tau) - K^{-1}f_{3-i}I(\tau)\right]\\ &= (-1)^i\sqrt{\frac{\mu_0}{\gamma t_1}}\left[\frac{\chi_1-\chi_2}{2}b_i - K^{-1}f_{3-i}\right]I(\tau).\end{aligned} \tag{22}$$

Passing to dimension physical quantities, we find from (9), (21), and (22) expressions for the tangential

components of **E** at points on the conductor surface Γ:

$$E_{\mathrm{i}}^{\Gamma} = \sqrt{\frac{\mu_0}{\pi\gamma}} b_{\mathrm{i}} \frac{\partial}{\partial t} \int_0^t \frac{I(\Theta)}{\sqrt{t-\Theta}} d\Theta + (-1)^i \left[\frac{k_1 - k_2}{2\gamma} b_{\mathrm{i}} - f_{3-\mathrm{i}} \right] I(t) + \ldots. \tag{23}$$

The projection of **E** onto the vector of the linear current density $\boldsymbol{\delta}(t)$ is

$$E_{-}^{\Gamma} = \frac{\delta_1 E_1^{\Gamma} + \delta_2 E_2^{\Gamma}}{\delta} = \sqrt{\frac{\mu_0}{\pi\gamma}} b \frac{\partial}{\partial t} \int_0^t \frac{I(\Theta)}{\sqrt{t-\Theta}} d\Theta + \left[\frac{(k_2 - k_1)(b_1^2 - b_2^2)}{2\gamma b} + \frac{b_1 f_1 - b_2 f_2}{b} \right] I(t) + \ldots. \tag{24}$$

The projection of **E** onto the direction perpendicular to $\boldsymbol{\delta}(t)$ is

$$E_{\perp}^{\Gamma} = \left[\frac{(k_1 - k_2) b_1 b_2}{\gamma b} - \frac{b_1 f_2 + b_2 f_1}{b} \right] I(t) + \ldots.$$

Note that the first term in (23), which is the first approximation to the asymptotic expansion of **E**, is an analog of impedance boundary conditions for a pulsed electromagnetic field. Vectorially, these conditions are written as

$$\mathbf{E} \approx \frac{1}{\sqrt{\pi\mu_0\gamma}} \frac{\partial}{\partial t} \left[\int_0^t \frac{\mathbf{B}(\Theta) \times \mathbf{n}}{\sqrt{t-\Theta}} d\Theta \right].$$

This formula is a solution to the Abelian integral equation [13]

$$\mathbf{B}(t) \times \mathbf{n} = \sqrt{\frac{\mu_0\gamma}{\pi}} \int_0^t \frac{\mathbf{E}(\Theta)}{\sqrt{t-\Theta}} d\Theta. \tag{25}$$

For a 1D field, a formula similar to (25) was derived in [3, 14]. In the latter work, a relationship between the field vectors at the metal boundary was deduced by applying the Duhamel integral to the problem of the penetration of a step magnetic field pulse into a conducting half-space. In this problem, the field distribution over space coordinate and time is found with the error function.

For the calculation of the **E** components on a conducting surface in cases admitting an exact solution, see Appendices 1–3.

CONCLUSION

If the skin depth is small, the electromagnetic field in the conductors and dielectric environment can be found by the boundary layer method. The limiting solution corresponds to the external field distribution for the case when the conductivity of the conductor is indefinitely high. The first and next approximations are found from joining conditions for the solutions at the boundary (conductor surface) and depend on the limiting value of the linear current density and the conductor geometry.

The first approximation to the field is found from joining conditions for the limiting external field on the conductor surface and corresponds to impedance boundary conditions. The second-order correction to the external field is defined by its disturbance arising from the normal component of the second-approximation to the magnetic induction of the field in the metal. It can be calculated as a solution to the Neumann problem in the environment. The second-order correction to the inner field (in the metal) must join with the second-order correction to the external field on the conductor surface.

In the asymptotic expansion of the electric field strength on the conductor surface, the first-order correction corresponds to impedance boundary conditions for the limiting magnetic field and varies with time like the Abelian integral of the derivative of the total current in conductors. The second-order correction to the electric field strength on the conductor surface depends on the geometry of this surface (on its principle curvatures) and varies with time like the electric current in the conductors.

APPENDIX 1

Skin Effect in a Circular Conductor

A variable electric field in a circular conductor is described by the equations

$$\frac{d^2E}{dr^2} + \frac{1}{r}\frac{dE}{dr} - j\omega\mu_0\gamma E = 0, \quad \left.\frac{dE}{dr}\right|_{r=R} = -j\omega\mu_0\delta, \quad |E|_{r=0} < \infty, \tag{A1.1}$$

where ω is the alternating current frequency, R is the radius of the conductor, and δ is the linear current density when the conductivity of the conductor is indefinitely high.

A solution to this problem can be expressed through Bessel functions, which, in turn, can be represented in terms of asymptotic expansions with the proviso $(\omega\mu_0\gamma)^{1/2}r \longrightarrow \infty$. Thus, basically one can obtain the expansion of E in powers of a small parameter and compare it with (9), (19), and (20).

A simpler way is to seek a solution to problem (A1.1) in the vicinity of $r = R$ as an expansion in powers of $\varepsilon = (\omega\mu_0\gamma R^2)^{-1/2}$. To do this, we change variables $r =$

$R(1 - \varepsilon x)$ in (A1.1) to obtain

$$\frac{d^2E}{dx^2} - \frac{\varepsilon}{1-\varepsilon x}\frac{dE}{dx} - \beta^2 E = 0,$$
$$\left.\frac{dE}{dx}\right|_{x=0} = \beta^2 A, \quad E|_{x\to\infty} \longrightarrow 0, \tag{A1.2}$$

where

$$A = \sqrt{\frac{\omega\mu_0}{\gamma}}\delta, \quad \beta = \sqrt{j}.$$

Note that this value of β is equivalent to those used in (14) and other formulas in the text, since we can consider that the characteristic time is $t_1 = \omega^{-1}$ for alternating current.

We seek a solution to (A1.2) in the form of the series $E = E_0 + \varepsilon E_1 + \varepsilon^2 E_2 \ldots$. Substituting this series into (A1.2) and equating the coefficients at equal powers of ε, we come to a set of boundary-value problems for the ordinary differential equations

$$\frac{d^2E_0}{dx^2} - \beta^2 E_0 = 0, \ \left.\frac{dE_0}{dx}\right|_{x=0} = \beta^2 A, \quad E_0|_{x\to\infty} \longrightarrow 0,$$

$$\frac{d^2E_1}{dx^2} - \beta^2 E_1 = \frac{dE_0}{dx}, \quad \left.\frac{dE_1}{dx}\right|_{x=0} = 0, \quad E_1|_{x\to\infty} \longrightarrow 0.$$

Hence,

$$E_0^\Gamma = \beta\sqrt{\frac{\omega\mu_0}{\gamma}}\delta, \tag{A1.3}$$

$$E_1^\Gamma = -\frac{1}{2}\sqrt{\frac{\omega\mu_0}{\gamma}}\delta. \tag{A1.4}$$

Formulas (A1.3) and (19) totally coincide. Formula (A1.4) is also coincident with (20), because the external field does not influence the current distribution inside the conductor in our case ($\hat{B}_{i1}^\Gamma = 0$) and the curvatures are, respectively, $\chi_1 = k_1 R = 1$ and $\chi_2 = k_2 R = 0$.

APPENDIX 2

Skin Effect in a Conducting Ball

Consider an electromagnetic field generated by alternating current I with a frequency ω that passes in a conducting ball of radius R. The ball is connected to an ac source at the north and south poles. On the conducting space boundary Γ, which is a sphere, the magnetic induction has only one component in the direction of the latitude: $B^\Gamma = B_q/\sin\Theta$, where $B_q = \mu_0 I/(2\pi R)$ is the induction on the equator and $\Theta \in [0, \pi]$ is the angular coordinate (latitude) of a point.

In this example, the radial component of the magnetic induction on the boundary Γ is absent; therefore, the distributions of the current and magnetic field inside the ball do not affect the external field in Ω_e. Hence, boundary conditions for second approximation (11) are uniform and the components of $B_{i,1}$ in the skin layer depend only on the second term in (18), which takes into account the curvature of Γ. Then, according to (22), the second-order component in the asymptotic expansion of $\mathbf{E}^\Gamma$ equals zero, $E_{i,1}^\Gamma = 0$, and the second terms on the right-hand side of (23) and (24) also equal zero. We will try to verify this result by other analytical approaches.

The distribution of B in a conducting ball ($r \le R$) is given by

$$\frac{\partial}{\partial x}\left(x^2\frac{\partial B}{\partial x}\right) + \frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}\left(\sin\Theta\frac{\partial B}{\partial\Theta}\right) - k^2x^2B = 0, \tag{A2.1}$$

where $k^2 = j\omega\mu_0\gamma R^2$ and $x = r/R$.

We seek a solution in the form of the sum

$$B(x, \Theta) = \sum_{n=0}^{\infty}\Psi_n(x)P_n(\cos\Theta), \tag{A2.2}$$

where $P_n(\cos\Theta)$ are Legendre polynomials.

Using the conventional method of separation of variables, we obtain from (A2.1) and (A2.2) a boundary-value problem for the ordinary differential equation

$$\frac{d}{dx}\left(x^2\frac{d\Psi_n}{dx}\right) - [k^2x^2 - n(n+1)]\Psi_n = 0,$$
$$\Psi_n|_{x=1} = \Psi_n(1), \quad |\Psi_n|_{x=0} < \infty, \tag{A2.3}$$

where $\Psi_n(1)$ is defined by expanding B^Γ in Legendre polynomials:

$$\sum_{n=0}^{\infty}\Psi_n(1)P_n(\cos\Theta) = \frac{B_q}{\sin\Theta}. \tag{A2.4}$$

A solution to problem (A2.3) can be expressed through Bessel function of order $n + 1/2$, which, in turn, can be represented as asymptotic expansions at $|kx| \longrightarrow 0$. Thus, basically the result desired can be obtained.

A simpler way is to seek a solution to problem (A2.3) in the vicinity of $x = 1$ in the form of an expansion in powers of the small parameter $\varepsilon = |k|^{-1} = (\omega\mu_0\gamma R^2)^{-1/2}$. Applying the conventional technique of constructing an asymptotic solution of boundary layer type to (A2.3), we get

$$\Psi_n(\xi) = \Psi_n(1)\exp(-\beta\xi)[1 + \varepsilon\xi + 0(\varepsilon^2)], \tag{A2.5}$$

where $\xi = (1 - x)/\varepsilon$, $\beta = \sqrt{j}$, $\mathrm{Re}B > 0$.

The electric field strength on the surface of the ball has only the meridional component

$$E^\Gamma = E_\Theta\big|_{r=1} = \frac{1}{\mu_0\gamma x}\frac{\partial(xB)}{\partial x}\bigg|_{x=1} = \frac{1}{\mu_0\gamma}\sum_{n=0}^{\infty}\Phi_n(1)P_n(\cos\Theta), \quad \text{(A2.6)}$$

where

$$\Phi_n(x) = -\frac{1}{x}\frac{d(x\Psi_n)}{dx}.$$

The function Φ_n can be found in the form of a series in powers of the small parameter ε from (A2.5) if the variable x is changed to $x = 1 - \varepsilon\xi$:

$$\Phi_n(\xi) = \Psi_n(1)\exp(-\beta\xi)\varepsilon^{-1} \times[-\beta - \varepsilon\beta\xi + 0(\varepsilon^2)]. \quad \text{(A2.7)}$$

The first two terms in brackets in (A2.7) do not depend on n. In view of this fact, substituting (A2.7) into (A2.6) yields

$$E^\Gamma = -\sqrt{\frac{j\omega}{\mu_0\gamma}}\frac{B_q}{\sin\Theta} + 0(\varepsilon^2) = \sqrt{\frac{j\omega\mu_0}{\gamma}}\delta(\Theta) + 0(\varepsilon^2), \quad \text{(A2.8)}$$

since $\delta(\Theta) = \mu_0^{-1}B_q/\sin\Theta$ is the linear current density for the perfect skin effect. Comparing (A2.8) with (9), (19), and (20), we find that the component of the second approximation $E^\Gamma_{i,1}$ in the asymptotic expansion of the electric field strength on the surface of the ball equals zero: $E^\Gamma_{i,1} = 0$.

This example demonstrates that the second-order correction in the asymptotic expansion of the electric field on the surface of conductors of specific geometry may equal zero.

APPENDIX 3

Skin Effect in a Nonuniform Electromagnetic Field Excited in a Conducting Half-Space

Consider an electromagnetic field generated by a fine single wire with an alternating current I of frequency ω. The boundary Γ of the conducting half-space ($x > 0$) is the plane $x = 0$. The wire runs parallel to this plane in the direction of the z axis at a distance a from it.

(i) **Exact solution.** The field in such a system can be described by the Helmholtz equation for the electric component of the field:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} - \alpha k^2 E = j\omega\mu_0 I\delta(x+a)\delta(y), \quad \text{(A3.1)}$$

where $k^2 = j\omega\mu_0\gamma$, δ is the delta function, E and $\partial E/\partial x$ are continuous functions in the vicinity of the boundary Γ, and

$$\alpha = \begin{cases} 1 & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x < 0. \end{cases}$$

An exact solution to problem (A3.1) can be found with the Fourier cosine transformation with respect to y:

$$\tilde{E}(x,s) = \frac{2}{\pi}\int_0^\infty E(x,y)\cos(sy)dy,$$

$$E(x,y) = \int_0^\infty \tilde{E}(x,s)\cos(sy)ds,$$

$$\frac{d^2\tilde{E}}{dx^2} - (s^2 + \alpha k^2)\tilde{E} = D\delta(x+a), \quad \tilde{E}\big|_{x\to\pm\infty} \longrightarrow 0, \quad \text{(A3.2)}$$

where $\tilde{E}(x, s)$ is the Fourier transform of the electric field strength, $\tilde{E}$ and $d\tilde{E}/dx$ are continuous functions at $x = 0$, and $D = j\omega\pi^{-1}\mu_0 I$.

When solving boundary-value problem (A3.2), one easily finds that

$$\tilde{E}(0,s) = \frac{D}{k^2}\exp(-sa)(s - \sqrt{s^2+k^2}) = -\frac{D}{k}\exp(-sa)\left(1 - \frac{s}{k} + \frac{s^2}{2k^2} - \ldots\right). \quad \text{(A3.3)}$$

Applying the inverse Fourier cosine transformation to the components of (A3.3) yields

$$E(0,y) = -\frac{\beta_1}{\pi}\sqrt{\frac{\mu_0}{\gamma}}\frac{a}{a^2+y^2}I + \frac{1}{\pi\gamma}\frac{a^2-y^2}{(a^2+y^2)^2}I - \frac{2a}{\pi\beta_1\sqrt{\mu_0\gamma^3}}\frac{(a^2-3y^2)}{(a^2+y^2)^3}I + \ldots, \quad \text{(A3.4)}$$

where $\beta_1 = \sqrt{j\omega}$ and $\mathrm{Re}\,\beta_1 \geq 0$.

(ii) **Solution with the boundary layer method.** In the approximation of perfect skin effect, the magnetic induction $\mathbf{B}_0$ in Ω_e and the linear current density δ on the boundary Γ can be found by the image method:

$$B_0^\Gamma = B_0(0,y) = \frac{\mu_0 aI}{\pi(a^2+y^2)}, \quad \delta = \frac{aI}{\pi(a^2+y^2)}.$$

Let us define a local system of coordinates (x_1, x_2, x_3) at points near the flat surface of the conductor. Let us have $x_1 = y/a$, $x_2 = z/a$, $x_3 = x/h_3$, $h_1 = h_{10} = h_2 = h_{20} = 1$, $t_1 = \omega^{-1}$, $K = a^{-1}$, $h_3 = (\omega\mu_0\gamma)^{-1/2}$, and $\varepsilon = h_3K = h_3/a$.

In the first approximation, the magnetic induction inside the conductor $B_{1,0}$ and the electric field on the surface $E^{\Gamma}_{2,0}$ are given, according to (14) and (19), by

$$B_{1,0} = \frac{\mu_0 I}{\pi a(1+x_1^2)}\exp(-\beta x_3),$$
$$E^{\Gamma}_{2,0} = \frac{1}{\pi a}\sqrt{\frac{\mu_0\omega}{\gamma}}\frac{\beta I}{(1+x_1^2)}, \quad \text{(A3.5)}$$

where $\beta = \sqrt{j}$ and $\mathrm{Re}\,\beta \geq 0$.

In accordance with (15), the normal component of the magnetic induction is

$$B^{\Gamma}_{3,1} = -\frac{2\mu_0 I}{\pi a\beta}\frac{x_1}{(1+x_1^2)^2} = -C\frac{y}{(a^2+y^2)^2},$$

where $C = 2\mu_0 a^2 I/(\pi\beta)$.

The value of $B^{\Gamma}_{3,1}$ found defines the Neumann boundary condition for the scalar potential U_1 in Ω_e in the second approximation:

$$\frac{\partial^2 U_1}{\partial x^2} + \frac{\partial^2 U_1}{\partial y^2} = 0, \quad x \leq 0, \quad -\infty < y < \infty,$$
$$\left.\frac{\partial U_1}{\partial x}\right|_{x=0} = -B^{\Gamma}_{31} = C\frac{y}{(a^2+y^2)^2}, \quad \text{(A3.6)}$$
$$U_1|_{x\to-\infty} \longrightarrow 0, \quad U_1|_{y\to\pm\infty} \longrightarrow 0.$$

To solve problem (A3.6), we apply the Fourier sine transformation with respect to y:

$$\tilde{U}(x,s) = \frac{2}{\pi C}\int_0^{\infty} U_1(x,y)\sin(sy)dy,$$
$$U_1 = C\int_0^{\infty}\tilde{U}_1(x,s)\sin(sy)ds.$$
$$\frac{d^2\tilde{U}_1}{dx^2} - s^2\tilde{U}_1 = 0, \quad x \leq 0,$$
$$\frac{d\tilde{U}_1}{dx} = \frac{s}{2a}\exp(-sa), \quad \tilde{U}_1|_{x\to-\infty} \longrightarrow 0.$$

Hence, $\tilde{U}_1(x,s) = (2a)^{-1}\exp[s(x-a)]$ and

$$U_1(x,y) = \frac{C}{2a}\frac{y}{(x-a)^2+y^2}.$$

Then, the tangential component of the second approximation of the external magnetic induction is

$$B_{y,1} = -\frac{\partial U_1}{\partial y} = -\frac{C[(x-a)^2-y^2]}{2a[(x-a)^2+y^2]^2}. \quad \text{(A3.7)}$$

Expression (A3.7) defines the boundary condition for the second approximation in terms of the boundary layer in Ω_i:

$$\frac{d^2 B_{1,1}}{dx_3^2} - \beta^2 B_{1,1} = 0, \quad x_3 \geq 0,$$
$$B_{1,1}|_{x_3=0} = B^{\Gamma}_{1,1} = -\frac{C(1-x_1^2)}{2a^3(1+x_1^2)^2}, \quad B_{1,1}|_{x_3\to\infty} \longrightarrow 0.$$
$$B_{1,1} = -\frac{\mu_0 I}{\pi\beta a}\frac{1-x_1^2}{(1+x_1^2)^2}\exp(-\beta x_3).$$

Hence, according to (13), we obtain

$$E^{\Gamma}_{2,1} = \frac{I}{\pi a\gamma h_3}\frac{1-x_1^2}{(1+x_1^2)^2}. \quad \text{(A3.8)}$$

Now, we substitute (A3.5) and (A3.8) into series (9) and pass to the physical coordinates (x, y, z). Eventually, we come to expansion (A3.4) restricted to the first two terms found by the boundary layer method.

This example demonstrates the effect of the external magnetic field nonuniformity on the second-order correction to the electric field strength on the surface of the conductor.

REFERENCES

1. B. É. Fridman and Ph. G. Rutberg, Prib. Tekh. Éksp., No. 2, 196 (2001) [Instrum. Exp. Techniques **44** (2) (2001)].
2. O. V. Tozoni and I. D. Mayergoyz, *Calculation of Three-Dimensional Electromagnetic Fields* (Tekhnika, Kiev, 1974).
3. G. A. Shneerson, *Fields and Transient Processes in Apparatus of Superhigh Currents* (Énergoatomizdat, Moscow, 1992; Nova Science, New York, 1997).
4. M. R. Ahmed, J. D. Lavers, and P. E. Burke, IEEE Trans. Magn. **24**, 138 (1988).
5. K. Ishubashi, IEEE Trans. Magn. **26**, 458 (1990).
6. S. M. Apollonskiĭ, *Calculation of Electromagnetic Shielding Shells* (Énergoizdat, Leningrad, 1982).
7. S. M. Rytov, Zh. Éksp. Teor. Fiz. **10**, 180 (1940).
8. S. Yuferev and N. Ida, IEEE Trans. Magn. **34**, 2605 (1998).
9. S. Yuferev and N. Ida, IEEE Trans. Magn. **35**, 1486 (1999).
10. P. L. Kalantarov and L. A. Tseĭtlin, *Inductance Calculation (Reference Book)* (Énergoatomizdat, Leningrad, 1986).
11. I. N. Vekua, *Foundations of Tensor Analysis and Theory of Covariants* (Nauka, Moscow, 1978).
12. J. Colle, *Perturbation Methods in Applied Mathematics* (Blaisdell Publishing Company, Waltham, 1968; Mir, Moscow, 1972).
13. V. I. Smirnov, *A Course of Higher Mathematics* (Nauka, Moscow, 1974; Addison-Wesley, Reading, 1964), Vol. 2.
14. V. M. Mikhaĭlov, *Pulsed Electromagnetic Fields* (Vishcha Shkola, Kharkov, 1979).

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1120–1124. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 52–56.

SOLIDS

Skin Effect in Massive Conductors Used in Pulsed Electrical Devices: II. Massive Conductors in Electric Circuits

B. E. Fridman

Institute of Problems in Electrophysics, Russian Academy of Sciences, St. Petersburg, 191186 Russia
e-mail: fridman@mail.infostar.ru

Received February 21, 2002

Abstract—Approximate relationships between the current and voltage drop across a set of massive conductors are established. They include two constant parameters: external inductance L of the conductors and skin parameter S. Specific shapes of the voltage across the massive conductors are illustrated with oscillograms. Mathematical processing of current and voltage oscillograms to estimate L and S is developed. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In [1], an electromagnetic field generated by massive conductors with current was analyzed and asymptotic relationships for the field vectors on the conductor surface were derived for a small skin depth. This work extends the approach used in [1] for massive conductors as elements of the electric circuits of pulsed devices.

Massive conductors withstanding currents as high as several megaamperes or several tens of megaamperes are a basic component of devices generating high current pulses [2]. Such conductors usually operate under conditions of small skin depth, and the dependence of the current on the voltage drop across them cannot, as a rule, be adequately described with a set of equivalent inductances and resistances.

The problem of electrical circuits with massive conductors is not new. In publications dealing with the subject, the pulsed current and voltage across massive conductors are usually determined with Laplace or Fourier transformations with respect to time with subsequent calculation of the transient operator resistance of the conductors. As a rule, conductor geometries used in previous works allow for the mathematical description of the magnetic field in 1D form. For given electric circuits, final analytical expressions for the current and voltage involve integrals of inverse Laplace (Fourier) transformation, which makes them approximate and sometimes rather awkward [3, 4].

In spite of a substantial number of previous publications, the problem of deriving general relationships between the current and voltage across massive conductors still remains urgent. This problem can be resolved for any geometry of the conductors in the case of the sharp surface effect, which is typical of sources of high current pulses and loads to which the pulses are applied. Asymptotic relationships between the electromagnetic field vectors on the surface [1] admit generalization in the form of a formula with current- and time-independent parameters. Similar to Ohm's law, this formula relates the current in massive conductors to the voltage drop across them. The aim of this work is to substantiate the use of these parameters and to develop oscilloscopic techniques for their measurement.

We will follow designations adopted in [1].

SKIN PARAMETER OF MASSIVE CONDUCTORS

In the case of the sharp surface effect, the voltage drop across a subcircuit with massive conductors can be divided by convention into two parts:

$$U = U_L + U_S,$$

where

$$U_L = L\frac{dI}{dt}$$

is the voltage component due to the self-inductance emf of the external field and

$$U_S = \int_\Lambda \mathbf{E}\cdot\mathbf{dl}$$

is the voltage component due to field penetration into the metal (the integral is taken along a line Λ on the conductor surface Γ that coincides with any of lines of the linear current density vector $\boldsymbol{\delta}$).

The dependence of U_S on the total current $I(t)$ is given in [1] (formulas (24) and (4)):

$$U_S(t) = S\frac{d}{dt}\int_0^t \frac{I(\Theta)}{\sqrt{t-\Theta}}d\Theta + R_S I(t) + \ldots. \qquad (1)$$

1063-7842/02/4709-1120$22.00 © 2002 MAIK "Nauka/Interperiodica"

Here,

$$S = \sqrt{\frac{\mu_0}{\pi\gamma}}\int_\Lambda \mathbf{b}(\mathbf{r})\cdot\mathbf{dl}$$

is a current- and time-independent parameter, which is suggested to be referred to as the skin parameter (skin resistance), and

$$R_S = \int_\Lambda\left[\frac{(k_2-k_1)(b_1^2-b_2^2)}{2\gamma b}+\frac{b_1 f_1 - b_2 f_2}{b}\right]dl$$

is the coefficient at the second-order correction, which has the property of active resistance and depends on the external field nonuniformity and the principal curvatures of the surface Γ on the current line trajectory Λ.

As a rule, R_S is small for massive conductors in actual megaampere devices.

The internal ac resistance of massive conductors is related to the skin parameter as

$$Z_S(\omega) = \beta S + R_S + \ldots = (1+j)\sqrt{\frac{\omega}{2}}S + R_S + \ldots.$$

For extended parallel conductors, the parameter S can be calculated with the method proposed in [3] for calculating the internal ac resistance of massive extended conductors,

$$S = \frac{1}{\sqrt{\gamma\mu_0}}\frac{\partial L}{\partial n},$$

where $\partial L/\partial n$ is a change in the external inductance of parallel conductors when points on the conductor contour are displaced normally by equal distances.

MASSIVE CONDUCTORS IN ELECTRICAL CIRCUITS OF PULSE DEVICES

We consider massive conductors for which the second-order correction R_S in the asymptotic expansion of U_S is small. If $I(0) = 0$, which is always the case for pulsed devices,

$$U_S = S\frac{d}{dt}\int_0^t \frac{I(\Theta)}{\sqrt{t-\Theta}}d\Theta = S\int_0^t \frac{I'(\Theta)}{\sqrt{t-\Theta}}d\Theta, \qquad (2)$$

where $I'(t) = dI/dt$.

Considering relationship (2) as an Abelian integral equation [5], we obtain

$$I'(t) = \frac{1}{\pi S}\frac{d}{dt}\int_0^t \frac{U(\Theta)}{\sqrt{t-\Theta}}d\Theta$$

or

$$I(t) = \frac{1}{\pi S}\int_0^t \frac{U(\Theta)}{\sqrt{t-\Theta}}d\Theta. \qquad (3)$$

The voltage U_S is impossible to measure in practice. Therefore, we deal with the voltage drop across a set of massive conductors having an external inductance L:

$$U = LI' + S\int_0^t \frac{I'(\Theta)}{\sqrt{t-\Theta}}d\Theta. \qquad (4)$$

Given $U(t)$, L, and S, the current $I(t)$ is determined from a solution to the Volterra integral equation

$$I(t) = \mathcal{T}(t) - \lambda\int_0^t \frac{I(\Theta)}{\sqrt{t-\Theta}}d\Theta, \qquad (5)$$

where

$$\mathcal{T}(t) = \frac{1}{L}\int_0^t U(\Theta)d\Theta, \quad \lambda = \frac{S}{L}.$$

Equation (5) is obtained by integrating (4).

A solution to Eq. (5) can be found by the iteration scheme or by applying Fourier or Laplace transformation. In the latter case, the Laplace transform of the current,

$$\bar{I}(p) = \frac{\bar{U}(p)}{L(p+\lambda\sqrt{p})}$$

admits the definition of the current $I(t)$ by quadratures [6]:

$$I(t) = \frac{1}{L}\int_0^t U(\Theta)K(t-\Theta)d\Theta, \qquad (6)$$

where

$$\bar{U}(p) = \int_0^\infty U(t)\exp(-pt)dt$$

is the Laplace transform of the voltage $U(t)$,

$$K(x) = \exp(\lambda^2 x)[1-\mathrm{Erf}(\lambda\sqrt{x})],$$

and

$$\mathrm{Erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y \exp(-z^2)dz$$

is the error function.

Formulas (2) and (3) or (4) and (6) can be considered as a specific Ohm's law for massive conductors. They can be used for finding currents and voltages in branched circuits. In addition, these formulas can be

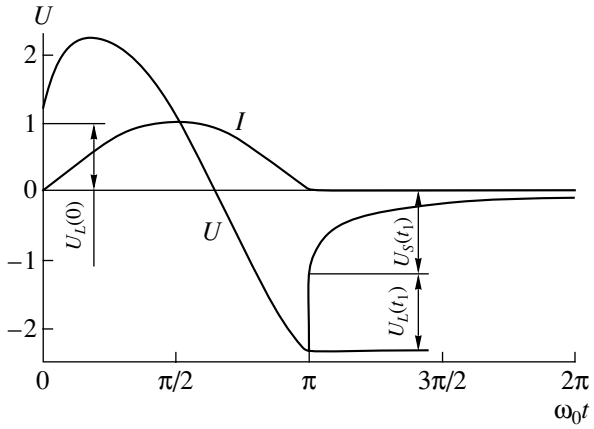


Fig. 1. Voltage U across a set of massive conductors for a half-period of a current pulse I. $I_m = 1$, $\omega_0 = 1$, $L = 1$, and $S = 1$.

written in terms of Kirchhoff's laws for multielement circuits of pulsed devices. Analysis of branched electric circuits with massive conductors is beyond the scope of our article and may be the subject of further investigation.

The voltage U across a massive conductor possesses specific features illustrated in Fig. 1 for a half-period sinusoidal current pulse

$$I(t) = \begin{cases} I_m \sin(\omega_0 t) & \text{for} \quad 0 \le t \le t_1 = \pi/\omega_0 \\ 0 & \text{for} \quad t \ge t_1 = \pi/\omega_0. \end{cases}$$

The plot $U(t)$ in Fig. 1 shows that the voltage changes stepwise initially, by an amount $U_L(0) = LI'(0)$, and at the end of the pulse, by an amount $U_L(t_1) = LI'(t_1)$. These voltage steps are explained by a voltage drop across the external inductance L of the set of conductors. Immediately after the current pulse, the voltage across the conductors equals

$$U_S(t_1) = S\int_0^{t_1} \frac{I'(\Theta)}{\sqrt{t_1 - \Theta}} d\Theta$$

and then asymptotically falls to zero, $U_S|_{t\to\infty} \longrightarrow 0$. Such a behavior results from the fact that, although the total current in the conductor equals zero, $I = 0$, partial currents may pass near the surface in opposite directions. In circuits with a low inductance L, the effect of nonzero residual voltage can be observed in oscillograms (Fig. 2).

MEASUREMENT OF SKIN PARAMETER IN MASSIVE CONDUCTORS

The skin parameter S, as well as the ohmic resistance, can be determined indirectly by measuring the current I and voltage drop U across a set of massive conductors. Basically, one can find S by measuring the ac resistance of massive conductors

$$Z(\omega) = j\omega L + \sqrt{j\omega} S$$

and by measuring the external inductance L separately at a high frequency (1 or 10 MHz). However, in the case of megaampere devices, the instruments cannot be connected in such a way that current lines coincide during operation and measurement. Therefore, S and L can be reliably measured for normally operating pulsed current sources and l loads by oscilloscopes measuring I and U. To suppress nonlinear effects due to high currents, the measurements should be made at a reduced pulse energy (5–20% of the typical value).

Figure 2 shows current (I) and voltage (U) oscillograms taken from massive conductors (buses) of an electrodynamic compression bench [7]. The buses were short-circuited by a copper rod 25 mm in diameter. The voltage oscillogram exhibits a residual potential after the end of the current pulse. This suggests that the skin parameter S comprises a significant portion of the conductor impedance.

The voltage U across a set of conductors can be represented in the form

$$U(t) = LI'(t) + RI(t) + S\frac{d}{dt}\int_0^t \frac{I(\Theta)}{\sqrt{t - \Theta}} d\Theta. \qquad (7)$$

Here, $U(t)$ and $I'(t)$ are discontinuous functions. Therefore, the formula

$$V(t) = LF_1(t) + RF_2(t) + SF_3(t), \qquad (8)$$

obtained by integrating (7), is more suitable for calculating the coefficients L, R, and S. In (8),

$$V(t) = \int_0^t U(\Theta) d\Theta,$$

$$F_1(t) = I(t),$$

$$F_2(t) = \int_0^t I(\Theta) d\Theta,$$

$$F_3(t) = \int_0^t \frac{I(\Theta)}{\sqrt{t - \Theta}} d\Theta.$$

The functions entering into (8) are calculated from current ($I(t)$) and voltage ($U(t)$) oscillograms. As a rule, an oscillogram contains no less than 512 readings, and the time between readings is much shorter than the characteristic time of signal variation. Therefore, the simplest methods of numerical integration, such as Simpson formulas, provide sufficient accuracy in calculating $V(t)$ and $F_2(t)$. The convolution singular integral $F_3(t)$ with a singularity at $t = \Theta$ can be taken with the numerical integration formula derived in the

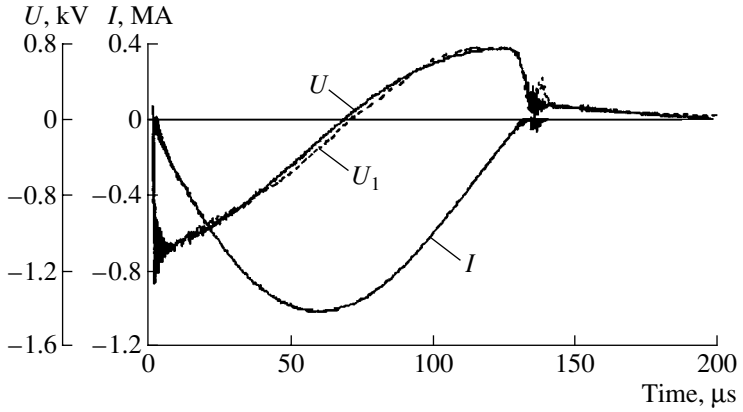


Fig. 2. Current (I) and voltage (U) oscillograms for the massive conductors (buses) of the electrodynamic compression bench [7]. U_1 is voltage calculated by (7) with $L = 28$ nH, $R = 0$, and $S = 1.93\ \Omega\ s^{1/2}$. The parameters were found by the least squares method from the current and voltage oscillograms. A bank of twelve capacitor modules (E7-25 capacitive storage) with a total capacitance of 25.9 mF was discharged at an initial voltage across the capacitors of 2.05 kV.

Appendix:

$$F_3(i\Delta t) = F_{3,i} \approx \sqrt{\Delta t}\left(\frac{4i-1}{2\sqrt{i}}I_i - \sum_{i'=1}^{i}\frac{I_i - I_{i'}}{\sqrt{i-i'}}\right); \quad (9)$$

$$i = 1, 2, \ldots,$$

where $I_i = I(i\Delta t)$ are readings on the current oscillogram $I(t)$.

The value of R can be evaluated from (8) with $t \longrightarrow \infty$. If the energy source is a capacitor bank of capacitance C, the resistance is calculated by the formula

$$R = \left.\frac{V(t)}{F_2(t)}\right|_{t\to\infty}\frac{1}{C[U_C(0) - U_C(\infty)]}\int_0^\infty U(t)dt, \quad (10)$$

where $U_C(0)$ and $U_C(\infty)$ are the voltages across the capacitors before and after the pulsed process.

The parameters L and S can be found from the voltage steps at the beginning and end of the current pulse. However, the measurements will be inaccurate, since the voltage steps under actual conditions are accompanied by high-frequency oscillations in both electric circuits and instruments, which makes it impossible to measure the steps accurately.

The accuracy in determining the parameters L and S will be the highest if one uses (totally or partially) arrays of current $I(t)$ and voltage $U(t)$ readings measured on the related oscillograms. The high accuracy is provided the least squares method. In this case, the parameters L, R, and S are found as a solution to a set of third-order linear algebraic equations

$$a_{i1}L + a_{i2}R + a_{i3}S = d_i; \quad i = 1, 2, 3, \quad (11)$$

where

$$a_{ii'} = \int_{t_0}^{t_f} F_i(t)F_{i'}(t)dt, \quad d_i = \int_{t_0}^{t_f} V(t)F_i(t)dt$$

$$(i, i' = 1, 2, 3),$$

and t_0 and t_f are the initial and final time instants on the oscillogram that are used for calculating the unknown coefficients.[1]

As t_0 and t_f, it is reasonable to take the instant of high-frequency oscillation disappearance and the end of oscillogram recording, respectively.

Figure 2 shows the current and voltage oscillograms taken when the capacitive energy storage was discharged into the buses of the electrodynamic compression bench [7]. From them, the parameters L, S, and R were found by the least squares method [Eqs. (11)] (see the caption to Fig. 2). This figure also shows the function U_1 calculated by formula (7) with L, R, and S found.

CONCLUSION

A new time-independent parameter (skin parameter) for characterizing massive conductors of pulsed devices in the case of the sharp surface effect is suggested. It is a specific analog of resistance in conventional Ohm's law in the sense that it relates the current and voltage drops across the massive conductors in the time domain.

The voltage across the massive conductors has a specific shape; namely, it is still retained after the current pulse and asymptotically decays with time. This

[1] If it is assumed that $R = 0$ or equality (10) is used to calculate R, the least squares method leads to a set of second-order linear equations for L and S.

feature substantiates the use of methods for analyzing transients in the time domain by invoking the skin parameter, since an infinitely large time of voltage existence across massive conductors is taken into account in a natural way in this case.

The skin parameter (skin resistance) and the external inductance can be measured indirectly by mathematically processing current and voltage oscillograms taken during the operation of pulsed devices.

APPENDIX

Calculation of Convolution Integral

$$F_3(t) = \int_0^t \frac{I(\Theta)}{\sqrt{t-\Theta}}\,d\Theta$$

The singularity in the integrand is removed with the transformation

$$F_3(t) = I(t)\int_0^t \frac{d\Theta}{\sqrt{t-\Theta}} - \int_0^t \frac{I(t)-I(\Theta)}{\sqrt{t-\Theta}}\,d\Theta \qquad \text{(A1)}$$
$$= 2\sqrt{t}I(t) - J(t).$$

The integral

$$J(t) = \int_0^t \frac{I(t)-I(\Theta)}{\sqrt{t-\Theta}}\,d\Theta$$

has no singularities and can be taken with the Simpson formula

$$J(i\Delta t) = J_i = \sqrt{\Delta t}\left(\frac{I_i - I_0}{2\sqrt{i}} - \sum_{i'=1}^{i-1}\frac{I_i - I_{i'}}{\sqrt{i-i'}}\right);$$
$$i = 1, 2, \ldots.$$

Substituting this equality into (A1) and taking into account that $I_0 = I(0) = 0$, we come to formula (9) for the function F_3.

REFERENCES

1. B. É. Fridman, Zh. Tekh. Fiz. **72** (9), 44 (2002) [Tech. Phys. **47**, 1112 (2002)].
2. B. É. Fridman and Ph. G. Rutberg, Prib. Tekh. Éksp., No. 2, 196 (2001) [Instrum. Exp. Techniques **44** (2) (2001)].
3. G. A. Shneerson, *Fields and Transient Processes in Apparatus of Superhigh Currents* (Énergoatomizdat, Moscow, 1992).
4. V. M. Mikhaĭlov, *Pulsed Electromagnetic Fields* (Vishcha Shkola, Kharkov, 1979).
5. V. I. Smirnov, *A Course of Higher Mathematics* (Nauka, Moscow, 1974; Addison-Wesley, Reading, 1964), Vol. 2.
6. *Tables of Integral Transforms (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1954; Nauka, Moscow, 1969), Vol. 1.
7. B. É. Fridman and Ph. G. Rutberg, Izv. Akad. Nauk, Énerg., No. 2, 23 (1998).

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1125–1129. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 57–62.

SOLIDS

Plastic Flow Localization upon Stretching Zr–1% Nb Alloy

T. M. Poletika, V. I. Danilov, G. N. Narimanova, O. V. Gimranova, and L. B. Zuev

Institute of Strength Physics and Materials Science, Siberian Division, Russian Academy of Sciences, Akademicheskiĭ pr. 2/1, Tomsk, 634021 Russia

Received November 26, 2001

Abstract—The stages in the plastic flow curve and the localization of plastic strains upon stretching Zr–1% Nb alloy are considered. The localization pattern is found to correlate with a strain hardening law upon plastic flow. Data for the dislocation structure in strain localization regions are reported. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Recent in-depth speckle interferometry studies of macrodeformation have revealed that the process is microscopically nonuniform at any stage of stressing [1, 2]. It has been found that plastic flow conditions in a given portion of the stress–strain curve strictly correlate with the space–time distribution of the strain tensor components. It turns out that the strain localization pattern can serve as a measure, for example, of the material moldability [3].

In this work, we elaborate upon our previous studies [3, 4] and investigate strain localization in Zr–1% Nb alloy [5]. The nature of deformation and its stages in hcp Zr–Nb alloys are as yet little understood. There is no agreement among researchers even regarding the nature of plastic deformation and mechanisms of strain hardening of pure zirconium [6]. Little has also been reported concerning the stages of the plastic flow curve for Zr–1% Nb alloy, as well as their correlation with deformation localization and the microstructure. These data, however, are of great importance for clarifying the evolution of the microstructure; for finding reasons for deformation localization at the macrolevel and, consequently, for the formation of fracture nuclei; and eventually for estimating the ductility margin of Zr–1% Nb alloy.

EXPERIMENTAL

The Zr–1% Nb alloy studied in this work consisted of recrystallized α-Zr grains (of mean size 5 μm) with uniformly distributed β-Nb precipitates up to 0.1 μm in size. The recrystallization procedure took place in four stages: (i) hot pressing in the α range, (ii) water quenching from the β range starting from 1050°C, (iii) annealing at 580°C for 3 h, and (iv) three cold rollings with two intermediate annealings at 580°C for 3 h.

Flat specimens with the working part measuring $42 \times 5 \times 2$ mm were stretched on an Instron-1185 testing machine with a clamp rate of 0.1 mm/min. Stress–strain diagrams were recorded by laser speckle interferometry, and simultaneously the fields of point displacement vectors $\mathbf{r}(x, y)$ [1] on the specimen surface were fixed. From the latter, the components of the plastic strain tensor $\beta = \nabla \cdot \mathbf{r}(x, y)$ were calculated. For simplicity, we hereafter will deal with only one component: the local elongation $\varepsilon_{xx} = \partial u/\partial x$ (u is the component of the vector $\mathbf{r}$ along the stretching axis x). Electron microscopy studies were carried out in an ÉMV-125 microscope. Thin foils were prepared by polishing in a 90% CH_3OH + 10% $HClO_4$ electrolyte at a temperature of –30°C.

STRESS–STRAIN CURVES

To make stress–strain σ–ε curve analysis more informative, we additionally studied *s*–*e* curves (where *s* and *e* are the true stress and strain, respectively) [7], as well as the strain dependences of the strain hardening coefficients $d\sigma/d\varepsilon(\varepsilon)$ and $ds/de(e)$. The alloy under study hardens in three stages: transition, linear, and parabolic. In Fig. 1, where the stress–strain curve and the strain dependence of the strain hardening coefficient are plotted in the *s*–*e* coordinates, the linear stage is within e = 1.2–1.8%; here, the hardening coefficient ds/de roughly equals 600 MPa. The linear stage is followed by the parabolic stage, where the hardening coefficient continuously decreases. Starting with a strain of 4%, the coefficient becomes very small and approaches zero. According to [8], this is an indication of the fourth stage of strain hardening.

The stages of plastic flow become more obvious when the curve is constructed in the coordinates $\ln(s - s_0)$–$\ln e$, where $s_0 = \sigma_{0,2}(1 + \varepsilon_{0,2})$ (Fig. 2). In these coordinates, the initial (nonlinear) portion of the curve reflects the so-called transition stage [8], which sets in after the yield point. This stage is not parabolic, although the hardening coefficient decreases. This stage is followed by the linear stage of strain hardening ($s \sim e$), occupying the range 1.2–1.8%.

1063-7842/02/4709-1125$22.00

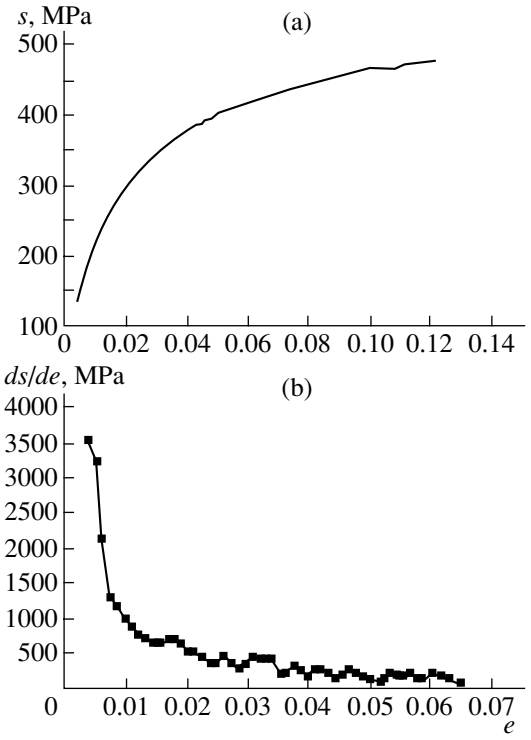


Fig. 1. (a) Plastic flow curve for Zr–1% Nb alloy in terms of true stresses and strains and (b) strain dependences of the strain hardening coefficient.

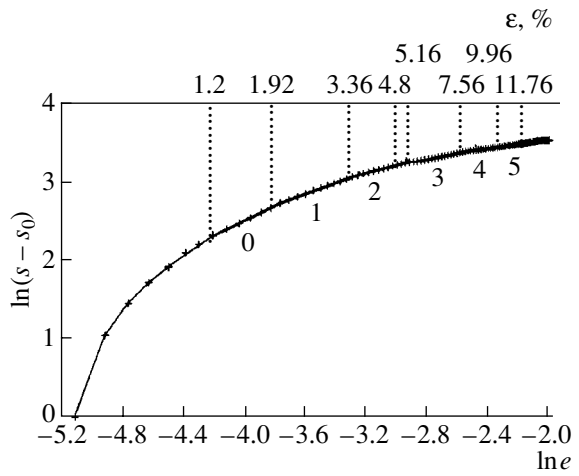


Fig. 2. Stages in the plastic flow curve.

Note that the parabolic stage of the stress–strain curve, in its turn, splits into five portions with the successively decreasing exponent of parabola $s \sim e^n$: n = 0.7, 0.5, 0.4, 0.3, and 0.2. The last portion of the parabolic portion with $n \longrightarrow 0$ manifests the formation of the neck, which is observed visually.

According to [6, 9], zirconium single crystals deform through prismatic slip and their stress–strain curve consists of three portions like that of better studied hcp Zn and Cd single crystals, whose orientation prevents basal slip [10]. The first stage of prismatic slip in Zr is characterized by the motion of screw dislocations with the Burgers vector $1/3\langle\bar{1}2\bar{1}0\rangle$. At the second stage of Zr hardening, screw dislocations gliding in two sets of prismatic planes interact. The third stage is the stage of dynamic recovery by means of cross slip from the prismatic to basal planes. When studying the strain hardening of Zr polycrystals, Akhtar and Teghtsoonian [9] also described the process by a parabolic function. They found that, when constructed in the $\sigma-\sqrt{\varepsilon}$ coordinates, the curve is divided into three straight portions with different slopes.

In the case of Zr–1% Nb alloy deformation, the situation is similar except that the number of stages increases to five. It can be supposed that, unlike pure Zr, basic factors responsible for plastic deformation in the alloy are the presence of matrix-incoherent β-Nb hardening disperse precipitates [5] and impurities such as oxygen, which are inevitably present in zirconium alloys [5, 11].

Both factors are of prime importance in the initial (nonparabolic) portion of the curve (the glide of screw dislocations over conjugate prismatic planes). At this stage, one can expect that the formation of dislocation loops and subsequently dislocation clusters around β-Nb particles may lead to strain hardening by the well-known Orovan mechanism [12]. A similar structure (dislocation clusters around Zr hydride precipitates) was observed in [13] at the linear stage of the stress–strain curve for Zr single crystals. This process is likely to develop at the linear stage in the strain range of 1.2–1.8%.

Impurities may also contribute greatly to strain hardening at the initial stage of plastic deformation. Here, oxygen-related effects [6] may become essential. For example, oxygen atoms may pin dislocations by the Cottrell atmosphere mechanism [12]. Another possibility is the stopping of dislocations by O–O pairs. The parabolic stage of hardening is apparently associated with cross slip to the Zr(0001) basal plane. This assumption is consistent with Bailey's data [13], who discovered basal slip near the hydride particles.

In all Zr–1% Nb specimens, the third stage began at a strain of ≈2%. Once basal slip has been initiated, the strain hardening coefficient starts decreasing. Long-term basal slip does not cause hardening [9, 10], hence, an extended portion of the stress–strain curve (4–12%) with a small exponent of parabola.

DEFORMATION LOCALIZATION PATTERNS

Laser speckle interferometry made it possible to visualize the distribution of ε_{xx} maxima at the various stages. At the very beginning of the curve (transition stage; $\varepsilon < 1.2\%$) and at the linear stage ($1.2\% < \varepsilon < 1.8\%$),

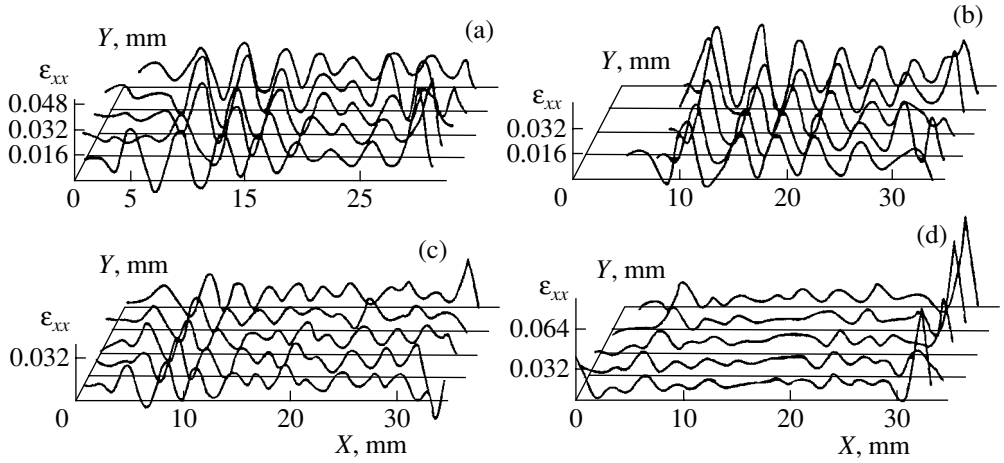


Fig. 3. Evolution of the strain localization pattern during stretching.

the local elongation maxima move. When the exponent of parabola is close to 0.7, a steady set of local elongation maxima arises (Fig. 3a) with a partial period $\lambda \approx 4.5$ mm.

Transition portions between the stages (nearly horizontal parts of the curve, $d\sigma/d\varepsilon = ds/de = 0$) deserve attention. Here, the distribution of local elongation (ε_{xx}) maxima changes: first, the spacing between the maxima shrinks roughly twofold (Fig. 3b), and then the maxima start moving, decreasing in height (Figs. 3c, 3d); that is, strain localization is absent. The same evolution of strain localization in the transition portions of the curve was observed in Al polycrystals with various grain sizes [14].

At the substage with $n < 0.5$, equidistant strain localization maxima begin to move until a neck forms, with the spacing between the localization fronts gradually decreasing. Figure 4 shows the position X of localization maxima in the specimen axis vs. total strain. The evolution of the strain localization pattern due to the change in the parabolicity exponent is clearly seen. Figure 5 demonstrates the dependence $\lambda(n)$, from which it follows that, in going from the linear stage to the parabolic one, λ first somewhat decreases, then ($0.3 < n < 0.5$) remains nearly constant, and finally ($n \approx 0.3$) decreases again.

Thus, the evolution of strain localization intimately correlates with the stages of the stress–strain curve. At the parabolic stages with $n \approx 0.7$ and 0.5, the steady set of strain localization maxima is observed. With $n < 0.4$, the localization fronts begin to move and the spacing between them diminishes.

Earlier [1, 2, 15], it was shown for various materials that strain localization fronts produce a steady set at the parabolic stage. In this work, a steady set of the maxima is found to occur only at $n \geq 0.5$ for Zr–1% Nb alloy. For $n < 0.5$, strain localization maxima move synchronously (λ is constant) up to $n \approx 0.3$. At small exponents n ($n < 0.3$), the motion loses synchronism and the maxima unite into a single localization peak, which then transforms into a microneck.

To conclude, the results obtained support our idea [16] that various forms of strain localization should be considered as self-organization in the imperfect structure of a deformed material [17].

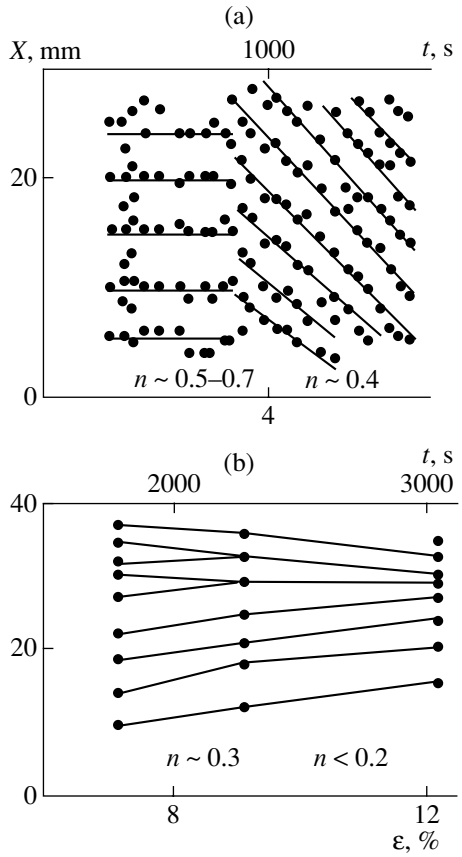


Fig. 4. Kinetics of strain localization: (a) initial and (b) final stage of the process.

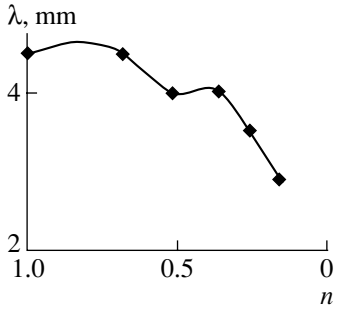
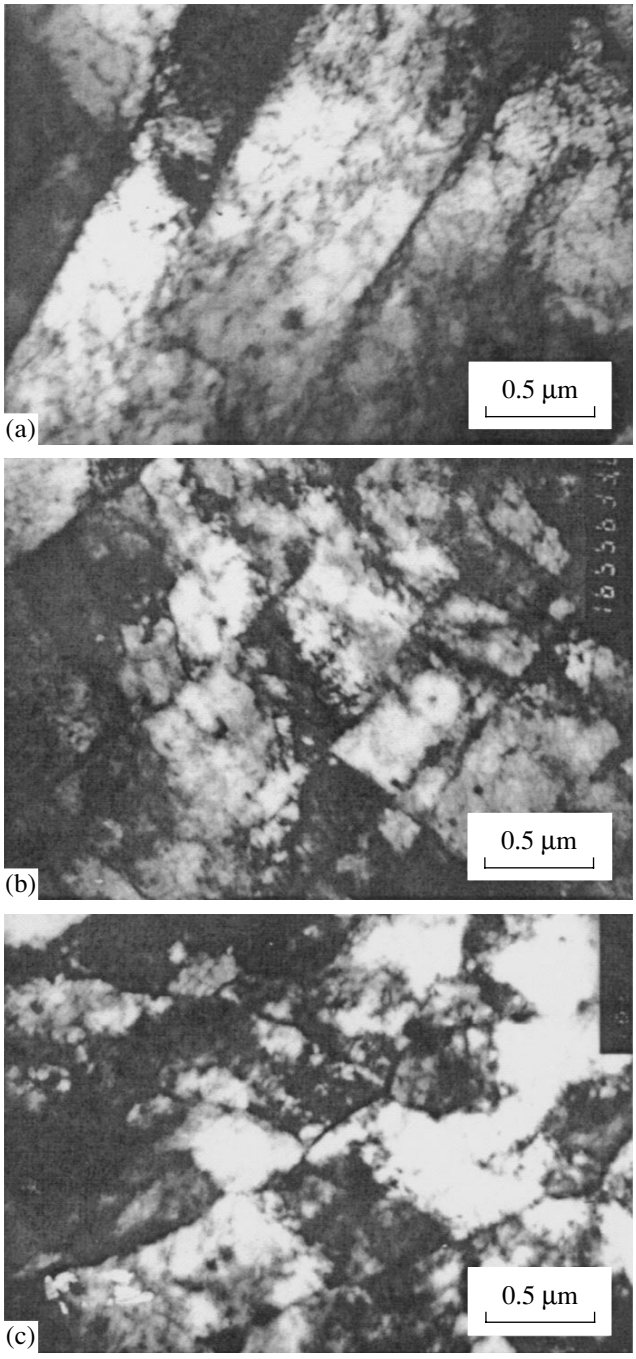


Fig. 5. Spacing λ between strain localization sites as a function of the exponent *n*.



DISLOCATION STRUCTURE OF THE DEFORMED ALLOY

Electron microscopy studies were performed on specimens strained to 7 and 12%, i.e., at those stages of the parabolic stress–strain curve where the strain localization fronts move and the spacing between them decreases. The banded substructure, which is typical of the substage with $n \approx 0.4$ (a total strain of 7%), is shown in Fig. 6a. It is seen that the bands are cut in the shear directions $\langle \bar{1}2\bar{1}0 \rangle$. The subboundaries of the bands are separated by a cellular–netlike dislocation substructure. Such a deformation pattern was observed in Zr single- and polycrystals at the third stage of strain hardening [6, 13]. The banded structure evolves into a 2D banded structure (Fig. 6b) at a strain of 12%. According to [8], the 2D structure appears because the number of locally active slip systems increases. It is known that gliding in Zr takes place largely over prismatic planes. If the strain (or temperature) is high, cross slip onto a basal plane is possible [6, 9]. In our dispersion-hardened alloy, cross slip is facilitated. This causes the 2D banded structure shown in Fig. 6b to form.

The microstructure of the alloy strained to 12% near the strain localization peak is depicted in Fig. 6c. The banded structure evolves further. Because of the interaction between variously directed subboundaries, the banded structure loses the clear-cut crystallographic orientation and splits into microbands that have a partially fragmented structure and a substructure with multidimensional discrete and continuous misorientations. Thus, for the parabolic substages with $n < 0.5$, first the 1D banded structure forms, which is followed by the 2D one (at $n \approx 0.4$) and the structure with multidimensional misorientations (at $n \leq 0.2$).

It is known that the formation of multidimensional banded structures and structures with multidimensional discrete and continuous misorientations takes place when dislocations of opposite sign annihilate. According to [8], this is typical of the fourth stage of fcc material deformation. A similar evolution of the dislocation structure is also observed in hcp materials, in particular, in Zr–1% Nb alloy hardened by disperse β-Nb particles.

The formation of the substructures typical of the fourth stage is accompanied by the motion of strain localization fronts and a decrease in the front spacing. Such an evolution of the component ε_{xx} at the fourth stage makes the strain distribution over the specimen more uniform. The high ductility of the material in this state is explained by successively involving its new constituents in the deformation process. That is why the alloy has a good cold ductility.

Fig. 6. Microstructures of the deformed Zr–1% Nb alloy.

CONCLUSIONS

(1) It is found that the stress–strain curve for dispersion-hardened Zr–1% Nb alloy is multiple-stage: it has a transition stage, as well as linear and parabolic stages of hardening. The last stage, in its turn, consists of five substages with different parabolicity exponents.

(2) The results of the laser speckle interferometry study suggest that the local elongation (ε_{xx}) distributions corresponding to the stages and substages are the following: (i) transition stage, origination and movement of strain localization sites; (ii) linear stage, movement of strain localization fronts; (iii) parabolic stage: steady set of strain localization fronts for $n < 0.5$, synchronous motion of strain localization fronts at $0.3 < n < 0.4$, and merging of the fronts into a single localization site with the resulting formation of a neck ($n < 0.3$).

(3) The electron microscopy examination of the alloy microstructure at the parabolic substages with $n < 0.4$ shows that the imperfect structure consists of 1D and 2D banded substructures, as well as substructures with multidimensional discrete and continuous misorientations. Such a dislocation structure is typical of the fourth stage of strain hardening.

REFERENCES

1. L. B. Zuev and V. I. Danilov, Philos. Mag. A **79** (1), 43 (1999).
2. L. B. Zuev, V. I. Danilov, and S. A. Barannikova, Int. J. Plast. **17** (1), 47 (2001).
3. L. B. Zuev, I. Yu. Zykov, V. I. Danilov, and S. Yu. Zavodchikov, Prikl. Mekh. Tekh. Fiz. **41** (6), 133 (2000).
4. T. M. Poletika, L. B. Zuev, and V. I. Danilov, Fiz. Met. Metalloved. **91** (5), 91 (2001).
5. E. Yu. Rivkin, B. S. Rodchenkov, and V. M. Filatov, *Strength of Zirconium Alloys* (Atomizdat, Moscow, 1974).
6. I. I. Papirov and G. F. Tikhinskiĭ, Preprint No. 76-23, KhFTI AN USSR (Kharkov, 1976).
7. Ya. B. Fridman, *Mechanical Properties of Metals* (Oborongiz, Moscow, 1952).
8. N. A. Koneva and É. V. Kozlov, *Structural Levels of Plastic Deformation and Destruction* (Nauka, Novosibirsk, 1990), pp. 123–186.
9. A. Akhtar and E. Teghtsoonian, Acta Metall. **19**, 655 (1971).
10. R. Berner and H. Kronmüller, *Plastische Verformung von Einkristallen*, in *Moderne Probleme der Metallphysik*, Ed. by A. Seeger (Springer-Verlag, Berlin, 1965; Mir, Moscow, 1969).
11. D. Douglass, *The Metallurgy of Zirconium* (IAEA, Vienna, 1971).
12. R. W. K. Honeycombe, *The Plastic Deformation of Metals* (Arnold, London, 1968; Mir, Moscow, 1972).
13. J. E. Bailey, J. Nucl. Mater. **7**, 300 (1962).
14. L. B. Zuev, B. S. Semukhin, and N. V. Zarikovskaya, Zh. Tekh. Fiz. **71** (5), 57 (2001) [Tech. Phys. **46**, 563 (2001)].
15. L. B. Zuev, S. A. Barannikova, and S. Yu. Zavodchikov, Fiz. Met. Metalloved. **87** (3), 77 (1999).
16. L. B. Zuev, Ann. Phys. (Leipzig) **10**, 965 (2001).
17. G. A. Malygin, Usp. Fiz. Nauk **169**, 979 (1999).

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1130–1133. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 63–66.

SOLIDS

Localization of Twinning Plastic Strain in Doped γ-Fe Single Crystals

S. A. Barannikova, V. I. Danilov, and L. B. Zuev

Institute of Strength Physics and Materials Science, Siberian Division, Russian Academy of Sciences, Akademicheskiĭ pr. 2/1, Tomsk, 634021 Russia

e-mail: levzuev@mail.tomsknet.ru

Received January 21, 2001; in final form, March 12, 2002

Abstract—The patterns of plastic flow localization in high-manganese γ-Fe fcc single crystals oriented for twinning upon stretching are obtained. Basic space–time features of strain localization at the stages of yield plateau, easy glide, and linear hardening are established. The velocity of strain localization sites during stretching is determined. Conditions under which plasticity autowaves appear in the strained medium are discussed. It is demonstrated that the local strain distributions in the case of twinning are similar to those due to dislocation glide. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Plastic deformation is known [1–3] to proceed locally over the entire range $\sigma_y < \sigma < \sigma_u$ (σ_y is the yield point and σ_u is the ultimate strength). The macroscopic forms of localization are few [2], and they are closely related to a strain hardening law, i.e., to the strain dependence of the hardening coefficient $d\sigma/d\varepsilon = \theta(\varepsilon)$. Most experimental data for the forms of localization are concerned with plastic deformation via dislocation glide. The no less important twinning mechanism [4] has received little attention [5, 6]. However, the character of plastic flow localization under the action of this mechanism, which is an alternative to dislocation glide, is of undeniable interest. In particular, since the localization of strain can be described in terms of an autowave process [1–3], it is necessary to realize whether such autowaves can form when plastic flow is of an essentially different nature.

EXPERIMENTAL

Samples used were single crystals of austenite high-manganese Hadfield (Fe–13wt % Mn–1.03wt % C) steel [7] grown by the Bridgman method in a helium atmosphere. An electrical discharge machine was used to cut out samples in the form of a double spade with a cross section of 1.6×5 mm and an operating length of 28 mm. The wide face, intended for observation, coincided with the (011) plane. A tensile stress was applied along the *x* axis aligned with the $[\bar{3}55]$ direction. The samples were homogenized for 24 h at 1373 K in an inert gas and quenched in water.

The samples thus prepared were stretched on an Instron-1185 testing machine at 300 K at a constant strain rate $\dot{\varepsilon} = 1.2 \times 10^{-4}$ s^{-1} (the velocity of the movable grip is $V_g = 0.2$ mm/min). The field of displacement vectors $\mathbf{r}(x, y)$ on the sample front surface was determined by speckle interferometry [8] in the range between the yield point and tensile strength at $\delta t = 36$ s intervals (increments in the total strain $\Delta\varepsilon_{tot} = \dot{\varepsilon}\,\delta t = 4 \times 10^{-3} = 0.4\%$). By numerical differentiation of $\mathbf{r}(x, y)$ with respect to coordinates *x* and *y*, the distributions of the longitudinal (ε_{xx}), transverse (ε_{yy}), shear (ε_{xy}), and rotational (ω_z) components of the plastic strain tensor $\beta_{ij} = \nabla r$ were calculated at each point of the surface under study. Then, for different time instants or various values of the total strain ε_{tot}, the local strain distribution over the whole sample or along its axial line was plotted. For this purpose, the local elongation in the stretching direction, $\varepsilon_{xx} = \partial u/\partial x$ (u is the r component along the stretching direction x), seems to be the most convenient component of the β_{ij} tensor. The space–time distributions $\varepsilon_{xx} = (x, y, t)$ will be correlated with different stages of the stress–strain curve $\sigma(\varepsilon)$.

PLASTIC FLOW AND STRAIN LOCALIZATION PATTERNS

Hadfield steel single crystals are especially convenient for correlating the shape of the stress–strain curve with the local strain distribution, since strain hardening parameters and the deformation mechanism depend on the crystallographic direction of stretching and the carbon content in γ solid solution in this case [9].

Because of the high deformation stress, which is due to the hardening of the solid solution by carbon, and the low stacking fault energy ($\approx 2.3 \times 10^{-2}$ J/m^2) in the γ-Fe single crystals studied, twinning induces strain immediately after the yield point [9]. The stretching of γ-Fe single crystals with a carbon content of ≈1% along the

[$\bar{1}11$] axis is known to result in compound twinning, while the stretching along the [$\bar{3}77$] axis results in ordinary twinning in the [$\bar{2}11$](111) system [9]. The intermediate [$\bar{3}55$] orientation lies on the same side of the standard crystallographic triangle. The flow curve of such crystals (Fig. 1) is therefore similar to that of the crystals oriented along the [$\bar{3}77$] axis; i.e., a sharp yield point and a yield plateau with $\theta = 0$ are observed. However, a relatively small (≈13°) deviation of the stretching axis from the [$\bar{1}11$] pole gives rise to stage I, where θ_I is small but nonzero. Finally, the activation of additional twinning systems and their interaction with each other and with rarely occurring acts of dislocation glide [5] increase to θ_{II} at stage II of the $\sigma(\varepsilon)$ curve.

All the features outlined above are observed in stress–strain curves *1–3* in Fig. 1. The sharp yield point is clearly seen. The following three stages of plastic flow are characterized by constant values of the strain hardening coefficient θ and are clearly defined in the θ–ε and θ–σ coordinates [10] (Figs. 2a, 2b): the yield plateau (the sharp yield point included) covers the range $0.7\% < \varepsilon_{tot} < 4\%$ with $\theta = 0$; stage I: $4\% < \varepsilon_{tot} < 10.9\%$ and $\theta_I \approx 337$ MPa $\approx 3 \times 10^{-3}$ G; and stage II: $11.6\% < \varepsilon_{tot} < 32\%$ and $\theta_{II} \approx 1045$ MPa $\approx 10^{-2}$ G.

In the range $11\% < \varepsilon_{tot} < 11.5\%$, stage I ends and stage II begins. The stage of parabolic hardening was observed in none of the crystals studied: they fail immediately after linear hardening.

We analyzed the local strain distributions by plotting the coordinates of the local strain peaks versus the stretching time t (at $\dot{\varepsilon} = \text{const} \cdot t \sim \varepsilon$) (Fig. 3). At the yield plateau stage, a single wide deformation zone, separating the strained and unstrained parts of the sample, starts propagating from the grip of the testing machine (Fig. 4). The strain is localized in this zone, while the rest of the sample remains almost unstrained. The velocity of the zone of local plastic flow was estimated from the slope of the plot in Fig. 3 at $V \approx 1.9 \times 10^{-5}$ m/s.

At the beginning of stage I, another strain front separates from this zone and propagates in the opposite direction over the strained part of the sample with a velocity of $V_I \approx -5.5 \times 10^{-5}$ m/s (Fig. 5). Prior to the separation, the velocity of the main zone, propagating in the unstrained part of the sample, decreases. This situation repeats twice during stage I (Fig. 3). The transition from stage I of easy glide to stage II is manifested by several local elongation peaks as the main strain zone passes to the grip (Fig. 3). At stage II, equispaced local-strain zones move synchronously with a constant velocity $V_{II} \approx -4.2 \times 10^{-5}$ m/s (Fig. 6). The spacing between the waves (wavelength) is $\lambda = (5 \pm 1) \times 10^{-3}$ m. The distributions of the local-strain zones here are more

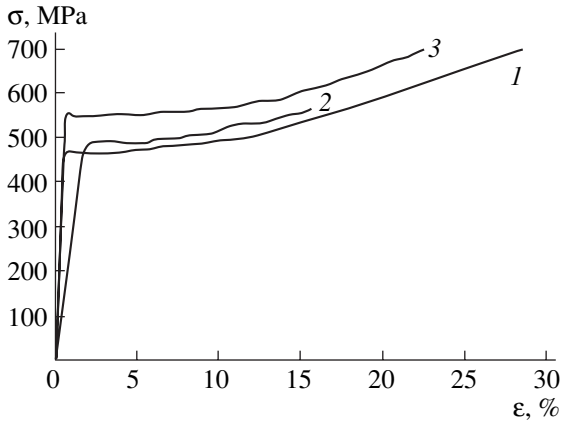


Fig. 1. Plastic flow curves for three Hadfield steel single crystals stretched along the [$\bar{3}55$] axis.

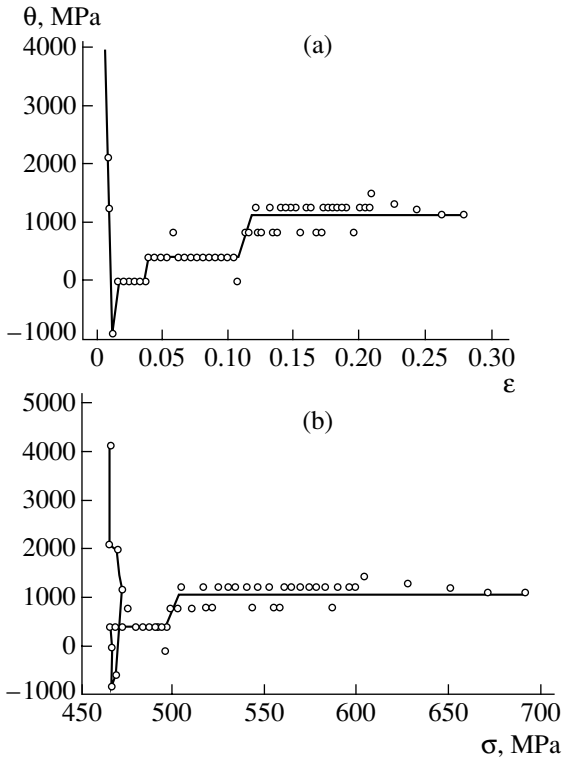


Fig. 2. Plastic flow stages in curve *1* (Fig. 1) in (a) θ–ε and (b) θ–σ coordinates.

complicated than those observed in previous studies. Stage II is characterized by the redistribution of the amplitudes of the local elongation peaks; however, the velocity and the wavelength of the strain localization remain constant.

When compared with the previous results [1–3, 11], our data allowed us to relate the stages in the stress–strain curve $\sigma(\varepsilon)$ to the space–time distributions of the strain tensor components. Our conclusions are as follows. (1) At the stage of yield plateau, deformation is due to the propagation of a single plastic front (Lüders

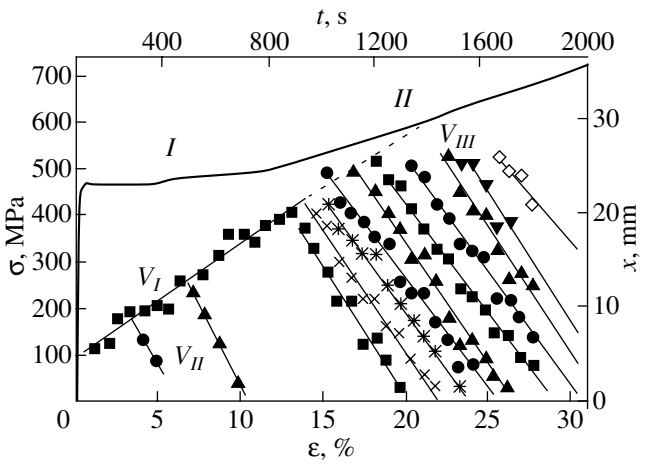


Fig. 3. Space-time evolution of the local strain peaks for the process illustrated by curve *1* in Fig. 1.

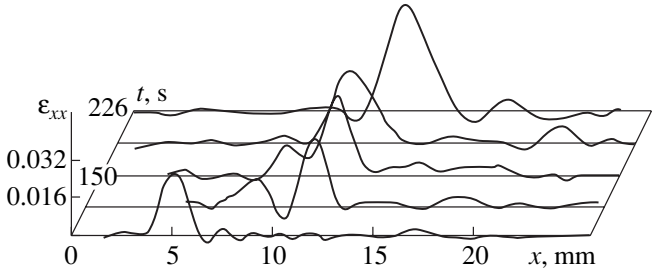


Fig. 4. Propagation of the initial solitary strain zone at the stage of yield plateau.

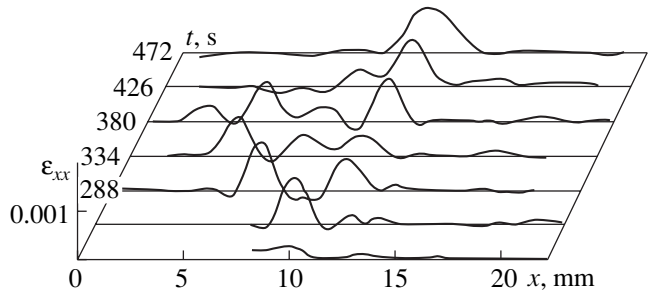


Fig. 5. Strain front separation at the stage of easy glide.

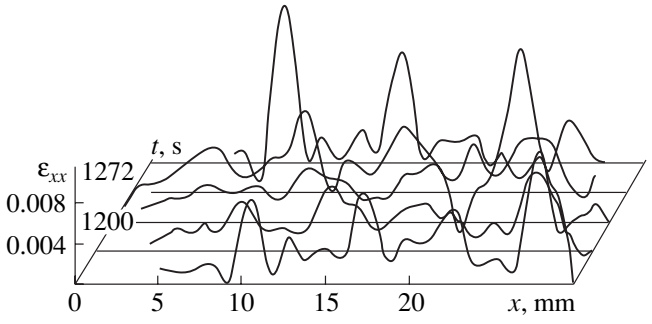


Fig. 6. Autowaves at the stage of linear hardening.

bands), which separates the elastic and plastic states of the material; (2) at stage I, such a wave gives rise to secondary waves (only one secondary wave can exist at each time instant), which propagate in the opposite direction in the strained medium; (3) at stage II, the generation of secondary waves becomes regular, so that a steadily moving sequence of equispaced plastic strain sites, i.e., a local-strain autowave, appears [3].

It should be noted that such a behavior of γ-Fe subjected to twinning deformation closely correlates with the behavior of single-crystal and polycrystalline metals and alloys subjected to dislocation glide deformation [2, 3, 11]. This means that the character of plastic strain localization is defined by the $\theta(\varepsilon)$ dependence rather than by a plastic flow microscopic mechanism.

DISCUSSION

The most important result of our study on plastic flow localization in Hadfield steel single crystals strained by twinning is the observation of plastic flow sites arising in the strained parts of the samples at stages I and II.

Figure 3 suggests the following scheme of the plastic flow process: once the lower yield point has been attained, a local plastic strain front starts propagating from one grip of the testing machine to the other with a constant velocity. Secondary fronts separate from the main one and move in the opposite direction also with a constant velocity. These fronts appear singly at stage I, while at stage II, a set of such parallel fronts forms a typical phase autowave with a wavelength $\lambda \approx 5 \times 10^{-3}$ m, a velocity $|V_{II}| \approx 4.2 \times 10^{-5}$ m/s, and frequency $f = V_{II}/\lambda \approx 8.4 \times 10^{-3}$ Hz. Thus, the initial site of plastic flow, i.e., the boundary between the strained and unstrained parts of the medium, generates the wave processes in the same manner as described in [12].

The sequence in which the secondary strain sites (local-strain zones) appear at stage I is of the most interest. It is known that, as twinning deformation occurs in one crystallographic system, the axis of tension precesses toward the $[\bar{1}11]$ pole, thus causing compound twinning. This is true if the strain is uniform over the sample. For a localized strain, the conditions required are met near the primary front, which slows down in this case. After the conjugate twinning system has come into action, an additional front arises, the stretching axis $[\bar{3}55]$ restores its initial direction, and the primary front velocity increases (Fig. 3). Additional sites develop in the prestrained zone and lead to material hardening. On the other hand, one may infer that the state of the medium changes drastically after the Lüders strain front has passed across it. Within the concept of an active medium where autowave processes are possible, this means a change in the type of an active medium [13]; for example, a medium that is bistable during the propagation of a single plasticity site

becomes self-oscillating at the stage of linear hardening.

CONCLUSION

Thus, the localization of plastic flow obeys much the same laws when plastic deformation is due to twinning and dislocation glide. The boundaries between the strained and unstrained parts are of key importance, since they may act as sources of autowaves of different types. It is worth noting that a similar situation with a pulse propagating in an inhomogeneous active medium was numerically studied in [14]. The solutions obtained in [14] relate the local characteristics of the active medium to the parameters of autowave processes, which are very close to the results of this study.

REFERENCES

1. L. B. Zuev, V. I. Danilov, and V. V. Gorbatenko, Zh. Tekh. Fiz. **65** (5), 91 (1995) [Tech. Phys. **40**, 456 (1995)].
2. L. B. Zuev and V. I. Danilov, Philos. Mag. A **79**, 43 (1999).
3. L. B. Zuev, Ann. Phys. (Leipzig) **10**, 965 (2001).
4. V. S. Boĭko, R. I. Garber, and A. M. Kosevich, *Reversible Plasticity of Crystals* (Nauka, Moscow, 1991).
5. P. Müllner and C. Solenthaler, Mater. Sci. Eng. A **230**, 107 (1997).
6. S. A. Barannikova, Zh. Tekh. Fiz. **70** (10), 138 (2000) [Tech. Phys. **45**, 1368 (2000)].
7. E. Houdremont, *Handbuch der Sonderstahlkunde* (Springer-Verlag, Berlin, 1956; Metallurgizdat, Moscow, 1959), Vols. 1, 2.
8. R. Jones and C. M. Wykes, *Holographic and Speckle Interferometry* (Cambridge Univ. Press, Cambridge, 1983; Mir, Moscow, 1986).
9. I. Karaman, H. Sehitoglu, Yu. Chumlyakov, *et al.*, Metall. Mater. Trans. A **32**, 695 (2001).
10. G. A. Malygin, Fiz. Tverd. Tela (St. Petersburg) **43**, 1832 (2001) [Phys. Solid State **43**, 1909 (2001)].
11. L. B. Zuev, V. I. Danilov, and S. A. Barannikova, Int. J. Plast. **17**, 47 (2001).
12. J. F. Bell, Int. J. Plast. **3**, 91 (1987).
13. V. A. Vasil'ev, Yu. M. Romanovskiĭ, and V. G. Yakhno, *Autowave Processes* (Nauka, Moscow, 1987).
14. A. N. Zaikin and T. Ya. Morozova, Biofizika **24** (1), 124 (1978).

Translated by A. Sidorova

Technical Physics, Vol. 47, No. 9, 2002, pp. 1134–1139. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 67–74.

SOLID-STATE ELECTRONICS

Hysteresis Loop Construction for the Metal–Semiconductor Phase Transition in Vanadium Dioxide Films

V. A. Klimov*, I. O. Timofeeva, S. D. Khanin**, E. B. Shadrin*, A. V. Ilinskiĭ***, and F. Silva-Andrade*****

**Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*
e-mail: shadr.solid@pop.ioffe.rssi.ru
***Gertsen State Pedagogical University, St. Petersburg, 191186 Russia*
****Autonomous University, Puebla, Mexico*

Received January 21, 2002

Abstract—Thermal hysteresis of the reflectivity of vanadium dioxide films observed upon the metal–semiconductor phase transition is studied. The major hysteresis loop is assumed to form when the phase equilibrium temperature in film grains and the grain size vary and correlate with each other. Within the suggested concept of hysteresis loop formation, it is demonstrated that the major loop may be asymmetric, i.e., broadened (shifted) toward lower temperatures. Unlike hysteresis branches for VO_2 bulk single crystal, those for VO_2 films are extended along the temperature axis and may exhibit a step if the grain size distribution has several maxima. The validity of the concept is verified experimentally. It is also shown that atomic force microscopy (AFM) data for the grain size distribution can serve to determine the distribution parameters from the phase equilibrium temperatures without constructing a complete set of minor hysteresis loops, as was required before. © *2002 MAIK "Nauka/Interperiodica".*

Vanadium dioxide, a material with metal–semiconductor phase transition, continues to attract widespread attention as a material for film interferometers used as a reversible holographic medium, optical limiters, temperature-sensitive elements, bolometers, etc. [1–3]. The devices listed exploit the fact that either the optical constants or the conductivity of vanadium dioxide changes stepwise as a result of the phase transition [4]. For example, a jump of optical constants of a VO_2 film placed between the mirrors of a film interferometer changes the optical path difference of the beams reflected from the mirrors, causing a jump in the reflectivity of the interferometer. The presence of thermal hysteresis allows for optical data storage. It is known [5] that optical memory devices operate reliably if the thermal hysteresis loop of the interferometer reflectivity is wide (15–20°C) and its branches are extended along the temperature axis as little as possible (4–6°C). The latter condition is necessary to provide the maximum modulation depth of the sensing radiation. In contrast, optical limiters require the thermal hysteresis loop to be as narrow as possible (3–5°C) with its branches extended along the temperature axis (25–35°C). This is necessary for the limiters to have a sufficient amount of linear negative feedback [2]. As follows from the aforesaid, the development of methods for controlling the parameters of the hysteresis loop during VO_2 film synthesis is of great practical interest. To do this, it is necessary to establish a correlation between the structure and physical properties of a VO_2 film used in the interferometer.

It is noteworthy that the thermal hysteresis loop of the reflection coefficient in an interferometer based on a VO_2 film is several times wider than that in an interferometer using a VO_2 single crystal (15 against 2°C). Moreover, it has extended (20–30°C) branches and is shifted toward lower temperatures (by 10–20°C). Reasons for such a difference have been discussed in [6]. In the work cited, hysteresis phenomena during the phase transition were treated in terms of the phase equilibrium temperature, T_{ci}, and coercive temperature distributions over microareas of the material. These distributions show to what extent one should recede from T_{ci} for the phase transition to occur intensely. According to [7], such microareas are film grains, and the deviation ΔT from the phase equilibrium temperature T_c at which a grain experiences the phase transition depends on the grain size [8]. The elementary hysteresis loop for each grain is "vertical" with respect to the temperature axis. This means that the reason for the difference in the hysteresis loop parameters for VO_2 films and bulk single crystals is the presence of a large number of grains of various size in VO_2 films. The grain size distribution in a 1080-Å-thick VO_2 film grown on a single crystal rutile substrate is visualized with the AFM image shown in Fig. 1. Note that the grains generally have not only different sizes but also different degrees of oxygen nonstoichiometry.

1063-7842/02/4709-1134$22.00 © 2002 MAIK "Nauka/Interperiodica"

The aim of this work was to elucidate factors responsible for the shape of the VO_2 hysteresis loop during VO_2 film synthesis. Preliminary studies have shown that our synthesis techniques cause a substantial spread in grain sizes and oxygen nonstoichiometry. This should be taken into account in treating experimental hysteresis data.

Let us consider the effect of a spread in the grain sizes more closely. We assume that each grain of the film has its own elementary hysteresis loop and each elementary loop is located normally to the temperature axis. The latter means that, at the transition temperature T^* for a specific grain, a nucleus of the metal phase grows through the semiconductor grain with the acoustic velocity [8]. A spread in the grain sizes causes a spread in the temperatures T^* of the transition to the metallic state (in the range from several degrees to several tens of degrees). The same is true for the temperatures T^{**} of the reverse transition to the semiconducting state (generally they vary in another range). The temperature spread is explained as follows. When the film is heated, the growth of a metallic nucleus in the semiconducting phase is stable if the temperature differs from the phase equilibrium temperature T_c by a value of ΔT_+ inversely proportional to the square root of the grain diameter (because of the martensitic character of the transition [7]). In the reverse transition, the stable growth of a semiconducting nucleus in the metallic phase takes place, as before, only if there is a temperature difference ΔT_- that also varies in inverse proportion with the square root of the grain diameter. In the general case, the proportionality coefficients in the dependence of the temperature deviation on the grain diameter are different for cooling and heating. Since the major hysteresis loop is the mere sum of the elementary loops of the grains, the distribution of the temperatures T^* and T^{**} in a wide interval widens the branches of the major loop of the film along the temperature axis compared with the vertically running branches of the major loop for bulk single-crystal VO_2. In addition, since the deviations ΔT_+ and ΔT_- from T_c for the film amount to several tens of degrees (against one or two degrees for the single crystal), the loop for the film is also much wider than for the single crystal.

Consider the formation of the major hysteresis loop for the film on a quantitative basis, i.e., by summing the elementary loops for grains. First, we assume that single-crystal grains of VO_2 differ in size rather than in oxygen nonstoichiometry. In this case, the elementary loops have equal positions on the temperature axis, $T_{ci} = T_c$, and the grain size distribution can be postulated or found from AFM images in the form of a function $\rho(d)$ (Fig. 1b). The histogram $\rho(d)$ makes it possible to find the dependence of the number of grains N vs. the deviation ΔT from T_c, $N(\Delta T)$, and determine the parameters of this function.

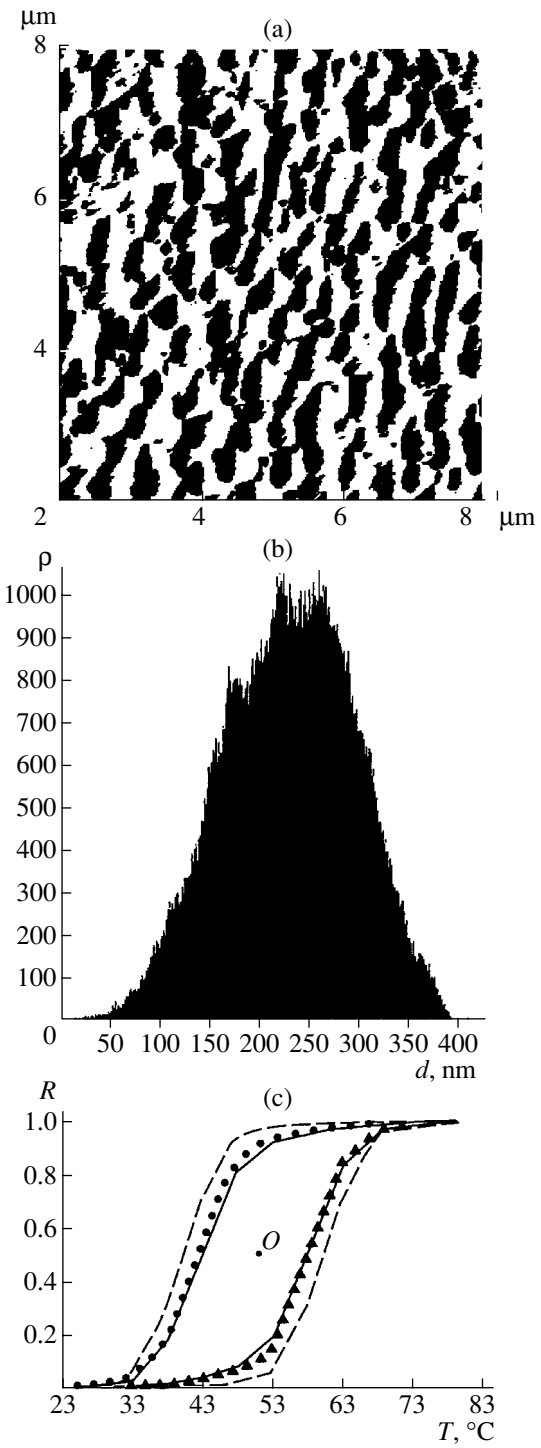


Fig. 1. (a) AFM image of the surface of the 1080-Å-thick vanadium dioxide film synthesized by laser ablation on a single-crystal rutile substrate, (b) distribution of grains of this film over sizes, and (c) normalized hysteresis loop for the reflectivity of a VO_2 film–based Fabry–Perot interferometer. In Fig. 1c: filled circles, data points; dashed line, calculation including only the grain size distribution; and solid curve, calculation including the distribution of grains over sizes and phase equilibrium temperatures. Experiment and calculation agree well if the grain distribution over phase equilibrium temperatures is Gaussian, the distribution maximum lies at 28.5°C, and the half-width of the distribution is 24°C.

The histogram in Fig. 1b depicts the simplest grain size distribution. Dependences $N(\Delta T_+)$ and $N(\Delta T_-)$ derived from such a distribution can be approximated by the Gaussian law: $N(\Delta T_+) = A_+\exp\{-(1/2)[(\Delta T_+ - T_{\max})/\Delta T_w]^2\}$ and $N(\Delta T_-) = A_-\exp\{-(1/2)[(\Delta T_- - T_{\max})/\Delta T_w]^2\}$, where $N(\Delta T)$ is the number of grains with a given temperature deviation. The meaning of the other parameters is clear from the definition of the Gaussian function.

To construct a mathematical model of the major hysteresis loop, it is necessary to know the grain size distributions $N^*(T^*)$ and $N^{**}(T^{**})$, where T^* and T^{**} are the temperatures of the semiconductor–metal and metal–semiconductor transitions. These functions cannot be found analytically, since any phase transitions beyond the hysteresis loop are absent. For the heating and cooling branches of the major loop, we have, respectively,

$$N(\Delta T_+) = N(T^* - T_c) = N^*(T^*)$$
$$= \begin{cases} 0 \quad \text{for} \quad -\infty < T^* < T_c, \\ A_+\exp[(-1/2)(T^* - T_{\max})^2/(\Delta T_{w-})^2] \\ \text{for} \quad T_c < T^* < +\infty, \end{cases} \tag{1a}$$

$$N(\Delta T_-) = N(T_c - T^{**}) = N^{**}(T^{**})$$
$$= \begin{cases} A_-\exp[(-1/2)(T^* - T_{\max})^2/(\Delta T_{w+})^2] \\ \text{for} \quad -\infty < T^{**} < T_c, \\ 0 \quad \text{for} \quad T_c < T^{**} < +\infty, \end{cases} \tag{1b}$$

where T_c is the phase equilibrium temperature, $T^*_{\max} = T_c + \Delta T_{+\max}$, $T^{**}_{\max} = T_c - \Delta T_{-\max}$, and A_+ and A_- are constants.

For the histogram in Fig. 1b, the heating and cooling branches of the major loop of the transition-induced jump ΔR in the reflectivity are expressed by the integrals

$$\Delta R_+(T) \sim \int_{-\infty}^{T} N^*(T^*)(D_+ + T^* - T_c)^{-6} dT^*$$
$$= \int_{T_c}^{T} N^*(T^*)(D_+ + T^* - T_c)^{-6} dT^*, \tag{2a}$$

$$\Delta R_c(T) \sim \int_{-\infty}^{T} N^{**}(T^{**})(D_- + T_c - T^{**})^{-6} dT^{**}. \tag{2b}$$

Note that integration in (2a) is over temperatures $T^* = T_c + \Delta T_+$ of the transition to the metallic state. The function $N(T^* - T_c) = N^*(T^*)$ represents the distribution of grains over the temperatures T^* of their transition to the metallic state. This function has a maximum $T^*_{\max} = T_c + \Delta T_{+\max}$. The additional weighting factors $(D + T^* - T_c)^{-6}$, where D = const, take into account the fact that the contribution of grains of a given size to the major loop is proportional to the total volume fraction of the new (metallic) phase inside the old one. In other words, this contribution depends not only on the number of grains of a given size but also on the metallic grain volume (i.e., on its mean diameter cubed). At the same time, the temperature deviations ΔT_+ and ΔT_- from the phase equilibrium temperature T_c vary inversely with the square root of the grain diameter [7]. When combined, these two dependences give the factors $(D_+ + T^* - T_c)^{-6}$ and $(D_- + T_c - T^{**})^{-6}$.

In (2b) (the cooling branch of the major loop), integration is over temperatures $T^{**} = T_c - \Delta T_-$ of the reverse phase transition to the semiconducting state. The function $N(T_c - T^{**}) = N^{**}(T^{**})$ has a maximum $T^{**}_{\max} = T_c - \Delta T_{-\max}$, which is shifted with respect to $T^*_{\max}$ by the major loop width. Here, one more assumption is made, namely, that the deviations ΔT_+ and ΔT_- upon heating and cooling are the same for all grains. This makes elementary loops symmetric about the phase equilibrium temperature T_c.

Both the cooling and heating branches of the hysteresis loop thus constructed have inclined temperature-dependent portions and also horizontal temperature-independent segments of length that is nearly equal to the half-width of the widest elementary loop (Fig. 1c, dashed line). The loop fits the experimental data well (Fig. 1c, data points) as regards its width and position on the temperature axis. However, the analytical and experimental shapes of the loop disagree. The discrepancy can be diminished by including the grain distribution over phase equilibrium temperatures T_c, as mentioned above.

In fact, grains in the film are usually nonstoichiometric to a variable extent in oxygen, which means that elementary loops are distributed not only over widths but also over phase equilibrium temperatures ($T_{ci} \neq T_c$). Oxygen nonstoichiometry stems from the fact that the equilibrium oxygen pressure over a grain depends on its radius of curvature. Because of this, the oxygen vacancy concentration varies from grain to grain in films synthesized. Oxygen vacancies behave like donors [9], causing a spread in phase equilibrium temperatures: the higher the free carrier concentration due to donors in a grain, the lower these temperatures. It should be noted that the dependence of the phase transition temperature on the free carrier concentration is a basic property of an electronic phase transition, which the metal–semiconductor transition in vanadium dioxide is, in our opinion [10]. Mathematically, a spread in temperatures T_{ci} can be included by introducing an additional (for example, Gaussian) grain distribution

over phase equilibrium temperatures under the integral sign:

$$\Delta R_+(T_c) \sim \int_{-\infty}^{T}\left[\int_{T_c}^{T} N^*(T^*)(D_+ + T^* - T_c)^{-6}dT^*\right] \quad (3a)$$
$$\times \exp[(-1/2)(T_c - T_{c\max})^2/(\Delta T_c)^2]dT_c,$$

$$\Delta R_-(T_c) \sim \int_{-\infty}^{T}\left[\int_{-\infty}^{T} N^{**}(T^{**})(D_- + T_c - T^{**})^{-6}dT^{**}\right] \quad (3b)$$
$$\times \exp[(-1/2)(T_c - T_{c\max})^2/(\Delta T_c)^2]dT_c.$$

Here, the second integration is over phase equilibrium temperatures T_c. The meaning of the constants $T_{c\,\max}$ and ΔT_c is obvious from the definition of the Gaussian function.

The grain distribution over T_c causes a fundamental change in the major loop shape: inclined portions of the branches appear in place of the horizontal segments. The branches of the major loop constructed with formulas (3) are shown in Fig. 1c. This loop is consistent with experimental data for the distribution parameters indicated in the caption to Fig. 1.

A more complex grain size distribution is exemplified in Fig. 2a, which is the AFM image of a (480 ± 80)-Å-thick VO_2 film synthesized on a quartz substrate by laser ablation, and in Fig. 2b, where the grain size distribution histogram is shown. As follows from the histogram, the distribution has at least two distinct peaks; namely, this film contains mostly 35- and 110-nm grains. This should be taken into consideration which constructing the major hysteresis loop. Figure 2c shows a hysteresis loop with step branches that is constructed according to (1) within our simple model, which ignores the spread in phase equilibrium temperatures between grains in this case. By comparing this analytical loop with that found experimentally by using formulas (3), one can extract information on the grain distribution over phase equilibrium temperatures, i.e., grain distribution over oxygen nonstoichiometry.

Preliminary conclusions are as follows. The simple model (grain distribution over sizes and oxygen nonstoichiometry) based on the findings in [6] allows us to explain why the major hysteresis loop of the reflection coefficient jump ΔR in the film is wider and shifts toward lower temperatures compared with that for bulk single-crystal VO_2 and why it has inclined portions instead of horizontal and vertical segments observed in the single crystals. At the same time, despite the asymmetry of its branches, the major loop remains symmetrical with the binary axis at the point *O* (Fig. 1c).

However, in many experiments [11], the hysteresis loop of the reflection coefficient for a vanadium dioxide interferometer was found to be asymmetric; that is, the heating and cooling branches ran in a different manner. Figure 3a shows the asymmetric loop of ΔR for an inter-

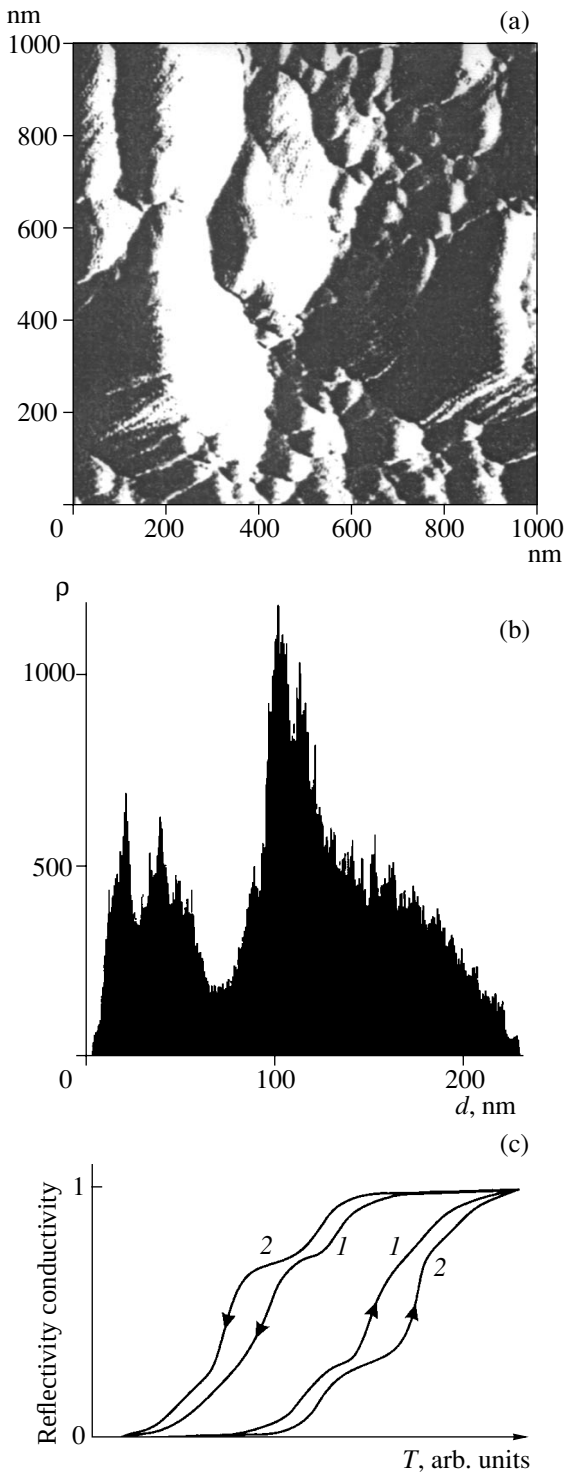


Fig. 2. (a) AFM image of the surface of the 480-Å-thick vanadium dioxide film synthesized by laser ablation on a single-crystal quartz substrate, (b) histogram of grain distribution over sizes, and (c) normalized thermal hysteresis loop (*1*) for the reflectivity of a VO_2 film–based Fabry–Perot interferometer constructed from the histogram in Fig. 2b on the assumption that a spread in phase equilibrium temperatures between grains is absent and (*2*) for the conductivity of this film. The histogram has two distinct maxima, as a result of which the theoretical and experimental hysteresis branches have a step form.

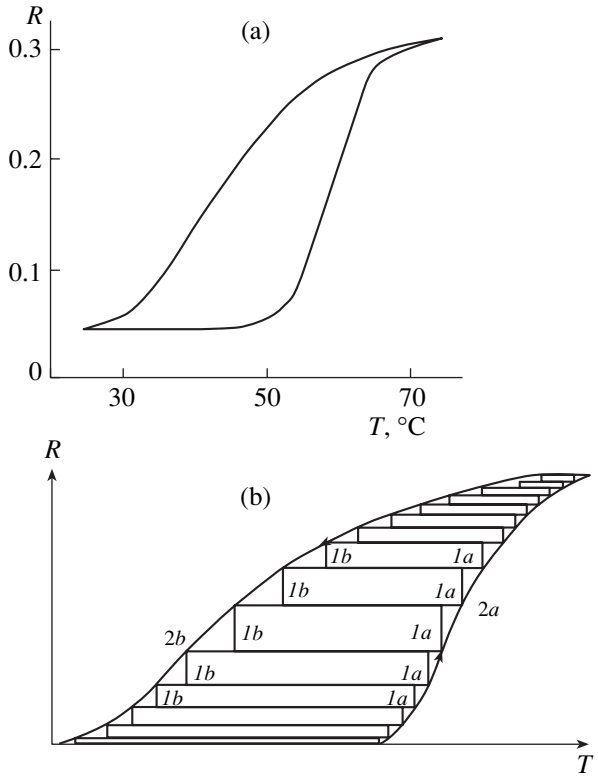


Fig. 3. (a) Experimental asymmetric hysteresis loop for the reflection coefficient and (b) the construction of the correlation asymmetry for this loop. In Fig. 3b, broken line *1* (*1a*, heating; *1b*, cooling) depicts the summation of several sets of elementary loops whose width ΔT is a linear function of the phase equilibrium temperature T_c (the base of the rectangles). The heights of the rectangles are proportional to the number of loops (defined by the Gaussian distribution) in a set. Curves *2a* and *2b* for heating and cooling are found from formulas (3a) and (3b), respectively, on the assumption that $\Delta T_+ = \Delta T_- = \Delta T$ are linear functions of T_c. Arrows indicate the temperature change direction.

ferometer with a 680-Å-thick vanadium dioxide film synthesized on a pyroceramic (Sitall) substrate. The film was synthesized by the electron-beam evaporation of metallic vanadium with subsequent annealing in oxygen under heavily nonequilibrium (oxygen-deficient) conditions.

Let us discuss various models allowing for the asymmetry of the major loop in this film.

The major loop constructed by summing elementary loops becomes asymmetric if the grain distributions over temperatures T^* of the semiconductor–metal transition (heating branch) and over temperatures T^{**} of the reverse transition (cooling branch) differ. A reason for such a difference may be a correlation between the grain size and oxygen nonstoichiometry. In other words, a correlation model suggests that a grain of a given size has a certain concentration of free carriers supplied by donors and thus a certain phase equilibrium temperature. Such an idea seems to be quite natural. Indeed, when VO_2 films are synthesized under conditions of oxygen deficiency, it is reasonable to suggest that fine grains have a higher concentration of oxygen vacancies (which behave like donors) and thus a lower T_c than coarse crystallites. The reason is that the equilibrium vapor pressure over a curved (liquid or solid) surface is always higher than over a flat surface; the larger the curvature, the larger the difference [12]. This means that oxygen atoms leave a curved surface more readily than a flat one, which leads to a higher saturation vapor pressure (recall that if we formally let the atom extraction energy approach infinity, the equilibrium vapor pressure will tend to zero [13]). Consequently, under oxygen-deficient conditions for VO_2 film synthesis, oxygen vacancies will arise primarily in grains with a low atom detachment energy (which require an increased vapor pressure to provide the stoichiometric composition), i.e., in fine grains (with the highest curvature). These grains, which have a lower phase equilibrium temperature, make a significant contribution to the low-temperature part of the major hysteresis loop. In fact, a small atom detachment energy in fine grains raises the probability of oxygen vacancy formation in them compared with coarse crystallites. The presence of the vacancies lowers the phase equilibrium temperature. It should be noted that the VO_2 film synthesis techniques applied in this work include the stage of oxidation of vanadium grains deposited on the substrate (resistive, electron-beam, and magnetron sputtering) or the oxidation of vanadium lower oxides (in laser sputtering) [14]. A correlation between grain sizes and degrees of nonstoichiometry ratios appears precisely at this stage.

In view of this correlation, formulas (3a) and (3b) should be integrated over elementary loops so that each successive term of the integral sum, which corresponds to a set of coarser grains with a higher phase equilibrium temperature, takes a temperature $T^* = T_c + \Delta T_+(T_c)$ of the semiconductor–metal transition that differs slightly from that for the preceding term (Fig. 3b, broken line *1a*). This makes the heating branch of the major loop steeper than the cooling branch. An increase in the temperature T_c turns out to be strongly compensated for by a decrease in the deviation ΔT_+ from T_c. As was indicated, this decrease is due to the correlation. At the same time, for each successive term of the integral sum, which corresponds to a set of coarser grains, the temperature $T^{**} = T_c - \Delta T_-(T_c)$ of the reverse metal–semiconductor transition shifts strongly toward higher temperatures. Such a significant shift is associated with both a decrease in the deviation ΔT_- from T_c and an increase in T_c (Fig. 3b, broken line *1b*). Because of this, the heating branch runs more smoothly than the cooling branch. When combined, both effects make the major loop asymmetric.

Mathematically, an expression for both branches of an asymmetric hysteresis loop can be derived as follows. Above it was assumed that the deviations ΔT_+ and ΔT_- from T_c for a given grain are the same: $\Delta T_+ = \Delta T_- =$

ΔT. Then, solving expressions $T^* = T_c + \Delta T_+(T_c)$ and $T^{**} = T_c - \Delta T(T_c)$ for T^* and T^{**}, respectively, and changing the variables in integrals (3a) and (3b), we reduce double integrals to repeated integrals. In this case, the integrands in formulas (3a) and (3b), which describe the heating and cooling branches, respectively, are different, although $\Delta T_+ = \Delta T_-$. As a result, the shapes of the heating, $\Delta R_+(T)$, and cooling, $\Delta R_-(T)$, branches of the major loop will be different; that is, the loop will be asymmetric.

Curves 2*a* and 2*b* in Fig. 3b are the branches of an asymmetric major loop constructed with formulas (3a) and (3b), respectively, with the proviso that ΔT_+ and ΔT_- are linear functions of T_c. A specific relationship between the width of an elementary loop and the phase equilibrium temperature for a grain is specified from physical considerations.

Comparing experimental asymmetric loops with those constructed by the approach suggested may provide important physical information about the mechanisms behind the action of intense particle fluxes, penetrating radiation, thermal or laser annealing, etc. on VO_2 films. These effects may disturb the stoichiometry ratio in the films and thus cause the asymmetry of the major hysteresis loop. These phenomena will be the subject of further investigation.

Thus, based on model assumptions (most of which were supported by direct AFM studies), we proposed a scheme for the formation of the major hysteresis loop of the reflection coefficient in a VO_2 film interferometer. In particular, it is assumed that a single new-phase nucleus originates in a grain. Then, it spontaneously grows and occupies the grain completely when the phase transition temperature is reached. Another assumption is that the nucleus grows in the grain at a fixed temperature, hence, the vertical run of elementary loops of the hysteresis branches. Finally, the most important feature of our approach is that a VO_2 film, unlike bulk single-crystal VO_2, consists of many grains, which are distributed over sizes and oxygen nonstoichiometry.

REFERENCES

1. A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, *Metal–Semiconductor Phase Transition and Its Application* (Nauka, Leningrad, 1979), p. 183.
2. V. P. Belousov, I. M. Belousova, O. B. Danilov, *et al.*, Proc. SPIE **3263**, 124 (1998).
3. V. Yu. Zernov, Yu. V. Kulikov, V. N. Leonov, *et al.*, Opt. Zh. **66** (5), 8 (1999) [J. Opt. Technol. **66**, 387 (1999)].
4. W. Bruckner, H. F. Opperman, W. Reichelt, *et al.*, in *Vanadiumdioxide* (Akademie, Berlin, 1983), p. 252.
5. V. L. Gal'perin, I. A. Khakhaev, F. A. Chudnovskiĭ, *et al.*, Zh. Tekh. Fiz. **61** (10), 194 (1991) [Sov. Phys. Tech. Phys. **36**, 1190 (1991)].
6. T. G. Lanskaya, I. A. Merkulov, and F. A. Chudnovskiĭ, Fiz. Tverd. Tela (Leningrad) **20**, 1201 (1978) [Sov. Phys. Solid State **20**, 193 (1978)].
7. I. A. Khakhaev, F. A. Chudnovskiĭ, and E. B. Shadrin, Fiz. Tverd. Tela (St. Petersburg) **36**, 1643 (1994) [Phys. Solid State **36**, 898 (1994)].
8. A. L. Roĭtburd, Usp. Fiz. Nauk **113**, 69 (1974) [Sov. Phys. Usp. **17**, 326 (1974)].
9. A. Leone, A. Trione, and F. Junga, IEEE Trans. Nucl. Sci. **37**, 1739 (1990).
10. A. V. Il'inskiĭ and E. B. Shadrin, Fiz. Tverd. Tela (St. Petersburg) **42**, 1092 (2000) [Phys. Solid State **42**, 1126 (2000)].
11. E. B. Shadrin, Doctoral Dissertation (St. Petersburg, 1997), p. 556.
12. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1995; Pergamon, Oxford, 1980).
13. R. Haase, *Thermodynamik der irreversiblen Prozesse* (D. Steinkopff, Darmstadt, 1963; Mir, Moscow, 1967).
14. V. N. Andreev, M. A. Gurvich, V. A. Klimov, *et al.*, Pis'ma Zh. Tekh. Fiz. **19** (9), 63 (1993) [Tech. Phys. Lett. **19**, 283 (1993)].

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1140–1144. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 75–79.

**OPTICS,
QUANTUM ELECTRONICS**

Laser-Induced Nonthermal Mechanical Stresses and Optical Damage in a Transparent Dielectric

V. N. Strekalov

Moscow State University of Technology (STANKIN), Moscow, 101471 Russia

e-mail: stvn@sec.ru

Received April 4, 2001

Abstract—Optical damage of pure transparent dielectrics in laser fields is considered. An elasticity theory problem concerning mechanical stresses that appear in the crystal lattice subjected to arbitrary spherically symmetric forces with a specified density is stated. A mathematically rigorous general solution to this problem is found, and arising boundary conditions are studied. A model of volume forces due to the action of hot nonequilibrium electrons on the lattice is described. In general, the early approximate results [1] are confirmed. The mechanism behind radiation-induced mechanical fracture of dielectrics proposed in this work is found to be adequate. Ways for the further study of this nonthermal inertia-free mechanism of optical damage, which has a purely mechanical nature, are discussed. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Optical damage to pure transparent dielectrics induced by focused laser radiation has been studied over many years. Associated results are embodied in many reviews and original papers. One of the latest reviews on this subject reports preliminary results of investigating a new nonthermal mechanism of optical damage [1].

Briefly, the damage mechanism [1] is as follows. An intense laser radiation causes the violent ionization of the material in the focal plane. As a result, a spatially inhomogeneous cloud of hot nonequilibrium electrons appears. The cloud swells the crystal, causing it to be damaged.

A more consistent description of the damage should take into account the momentum and energy fluxes transported by electrons from the center of the focal spot toward its periphery. When scattered by impurities or lattice imperfections, such fluxes produce volume forces, which strain the lattice. The strains are related to the mechanical stress tensor. The sample fails when the stresses exceed a threshold value (this is the natural way of mechanical fracture of a dielectric in a laser field). This process includes the formation of a primary crack, which is a macroinhomogeneity of the crystal. Electron scattering by the crack should be considered as the onset of the feedback mechanism, making the fracture process avalanche-like.

In this work, the stages of the primary crack formation will be given a detailed mathematical description.

It should be noted that fracture processes depend on experimental conditions, particularly, on radiation focusing. The distribution of the electron cloud parameters and the related momentum and energy fluxes can be considered as spherically symmetric if a short-focus lens is used for focusing. This case will be studied below. For a long-focus lens, a cylindrically symmetric model is more appropriate. It would be of interest to compare theoretical distinctions following from spherical and cylindrical models with the differences in associated experimental data. In particular, the shapes of the primary crack may differ.

For the spherically symmetric model, it is necessary to consider the radial stress tensor $\sigma_{rr} = \sigma_{rr}(r)$ and two coincident tangential tensors $\sigma_{\theta\theta} = \sigma_{\theta\theta}(r) = \sigma_{\varphi\varphi}(r)$. In what follows, it is shown that at certain distances r from the focal spot center, $\sigma_{rr} > 0$ (i.e., the lattice stretches), while in other regions, the lattice is compressed ($\sigma_{rr} < 0$). Under the conditions of our problem, the stresses $\sigma_{\theta\theta}(r)$ are always positive and stretch spherically symmetric layers of the sample. The sign of $\sigma_{\theta\theta}(r)$ is significant for the fracture of crystalline samples, since tensile strengths are approximately one order of magnitude lower than compression strengths.

Elaborating upon work [1], we will consistently state the physical problem in order to determine relative displacements and stress tensors. An equation for relative displacements and mechanical stresses will be solved in terms of the elasticity theory [2]. We will show that the problem can be solved exactly in the general form without invoking approximations used in [1]. A model of density of volume forces $\mathbf{F}(r)$ exerted by the electronic subsystem on the lattice in the spherically symmetric approximation will be discussed, and formulas enabling the numerical evaluation of arising mechanical stresses will be derived.

A solution will be found for the case of electron scattering by residual impurities in a pure transparent dielectric; however, the results will also be applicable

1063-7842/02/4709-1140$22.00

to other scattering modes and spherically symmetric forces $\mathbf{F}(r)$.

ELASTICITY EQUATION AND ITS SOLUTION

For a homogeneous elastic medium in the spherical coordinate system, one can introduce the relative displacements $u = u(r)$ of points in the medium, strain tensors $u_{rr} = u_{rr}(r)$, $u_{\theta\theta} = u_{\theta\theta}(r)$, and $u_{\varphi\varphi} = u_{\varphi\varphi}(r)$ (by virtue of symmetry, $u_{\theta\theta} = u_{\varphi\varphi}$), as well as mechanical stress tensors $\sigma_{rr} = \sigma_{rr}(r)$, $\sigma_{\theta\theta} = \sigma_{\theta\theta}(r)$, and $\sigma_{\varphi\varphi} = \sigma_{\varphi\varphi}(r)$ ($\sigma_{\theta\theta} = \sigma_{\varphi\varphi}$). Once forces acting on the medium are spherically symmetric and have a density $F(r)$, the function $u = u(r)$ must obey the equation from the elasticity theory (for details, see [2])

$$\frac{d}{dr}\left\{\frac{1}{r^2}\frac{d}{dr}[r^2u(r)]\right\} = -AF(r), \tag{1}$$

where

$$A = \frac{(1+\sigma)(1-2\sigma)}{E(1-\sigma)} \tag{2}$$

is a constant factor depending on the modulus of elasticity E and Poisson's ratio σ. Recall that σ usually varies from 0 to 0.5.

It should be emphasized that Eq. (1) is not the Laplace equation and contains the product $r^2u(r)$, unlike the latter. Therefore, singularities of the solution at the point $r = 0$ may appear. The point $r = 0$ deserves special mathematical study, although physically it is of no interest since the smallest values of the coordinate r that are physically justified depend on the lattice constant, Debye length, electron free path, and probably on other factors.

Integrating Eq. (1) twice yields its solution:

$$u(r) = C_1r + \frac{C}{r^2} - \frac{A}{r^2}\int\{r^2\int F(r)dr\}dr. \tag{3}$$

The quantities C_1 and C_2 are arbitrary constants of integration. It is important that one can add additional constants S_1 and S_2 to the integrals in (3). The values of S_1 and S_2 are taken for convenience (it is easy to check that S_1 and S_2 are merely added to C_1 and C_2 and have no influence on the form of the solution). With the constants S_1 and S_2, we can rearrange solution (3). To do this, we take into account that, according to the properties of integrals,

$$\int f(x)dx = \int_0^x f(x)dx + \text{const}, \tag{4}$$

where const is defined by the lower limit of integral (4).

In view of this property and letting c_1 and c_2 stand for combinations of the constants C_1, C_2, S_1, and S_2, we write solution (3) in the form

$$u(r) = c_1r + \frac{c_1}{r} - \frac{A}{r^2}\int_0^r\left\{r^2\int_0^r F(r)dr\right\}dr. \tag{5}$$

The introduction of the definite integrals makes it possible to numerically evaluate the values sought and is of considerable help in studying the behavior of the solution in the vicinity of zero.

Integrating the outer integral in (5) by parts yields

$$\int_0^r\left\{r^2\int_0^r F(r)dr\right\}dr = \frac{r^3}{3}\int_0^r F(r)dr - \frac{1}{3}\int_0^r r^3F(r)dr. \tag{6}$$

It turns out that the rearrangement performed simplifies appreciably the mathematical solution of the problem and allows for the exact solution for an arbitrary function $F(r)$.

Taking into account transformation (3), we can write the solution to (1) in the form

$$u(r) = c_1r + \frac{c_2}{r^2} - \frac{Ar}{3}I_1(r) + \frac{A}{3r^2}I_2(r), \tag{7}$$

where

$$I_1(r) = \int_0^r F(r)dr, \quad I_2(r) = \int_0^r r^3F(r)dr. \tag{8}$$

So far, no assumptions regarding the function $F(r)$ have been made, except that it must be spherically symmetric. Now, it is clear that it must be integrable so that integrals (8) be meaningful at any r. Any "physical" function meets this requirement.

BOUNDARY CONDITIONS

Boundary conditions for the problem under consideration must reflect two obvious physical requirements. First, the solution at zero should be either limited, $u(0) < 0$, or zero:

$$u(0) = \lim_{r\to 0} u(r) = 0. \tag{9}$$

The boundedness of the solution at zero would be a less strict requirement than condition (9). It might appear if condition (9) could not be met. Indeed, under certain conditions, requirement (9) is unsatisfiable; hence, $u(r) = u_0 > 0$. This result would mean that a vacuum cavity surrounded by a compressed high-density material forms at the center of the focal spot. The appearance of such a cavity in the focal region is an interesting physical phenomenon, which deserves further investigation.

The second boundary condition stems from the reasonable assumption that the radial component of the

mechanical stress must vanish at a point at infinity or on the free surface of the sample:

$$\sigma_{rr}(\infty) = \lim_{r\to\infty}\sigma_{rr}(r) = 0. \quad (10)$$

Boundary conditions (9) and (10) allow the unknown constants c_1 and c_2 to be determined.

Note that in the limit $r \longrightarrow 0$, we have for $I_1(r)$ and $I_2(r)$ in (7):

$$\lim_{r\to 0} I_1(r) = 0, \quad \lim_{r\to 0} I_2(r) = 0. \quad (11)$$

The vanishing of the second limit (11) implies that the limit of solution (7) itself at $r \longrightarrow 0$ has the form of 0/0 uncertainty. Therefore, to verify the fulfillment of boundary condition (9), the L'Hospital rule should be applied (a rather rare case for the solution of problems of mathematical physics). Separately differentiating the denominator r_2 and numerator $I_2(r)$ yields

$$\lim_{r\to 0} u(r) = \frac{A}{6}\lim_{r\to 0}\frac{1}{r}\frac{d}{dr}\int_0^r r^3F(r)dr = 0. \quad (12)$$

Thus, the first boundary condition can be met only by using the L'Hospital rule with $c_2 = 0$.

MECHANICAL STRESS TENSORS AND THE SECOND BOUNDARY CONDITION

The detailed derivation of formulas relating relative displacements with strain tensors and mechanical stress tensors was carried out in [2]. In the spherically symmetric system,

$$u_{rr}(r) = \frac{u(r)}{r}, \quad u_{\theta\theta}(r) = u_{\varphi\varphi}(r) = \frac{u(r)}{r} \quad (13)$$

and

$$\sigma_{rr}(r) = D[(1-\sigma)u_{rr} + 2\sigma_{\theta\theta}(r)], \quad (14)$$

$$\sigma_{\theta\theta}(r) = D[u_{\theta\theta}(r) + 2\sigma_{rr}(r)], \quad (15)$$

where

$$D = \frac{E}{(1+\sigma)(1-2\sigma)}. \quad (16)$$

For our problem, formulas (13)–(16) enable one to find the mechanical stress tensors:

$$\sigma_{rr}(r) = D\left[(1+\sigma)c_1 - \frac{1}{3}A(1+\sigma)I_1(r) - \frac{2A}{3r^3}(1-2\sigma)I_2(r)\right], \quad (17)$$

$$\sigma_{\theta\theta}(r) = D\left[(1+\sigma)c_1 - \frac{1}{3}A(1+\sigma)I_1(r) + \frac{A}{3r^3}(1-2\sigma)I_2(r)\right]. \quad (18)$$

Let the second boundary condition be fulfilled. Assuming that the integral $I_2(r)$ is meaningful, we find that

$$\lim_{r\to 0}\frac{1}{r^3}\int_0^r r^3F(r)dr = 0. \quad (19)$$

Then, for the second boundary condition to be fulfilled, the equality

$$c_1 = \frac{1}{3}AI_1(\infty) \quad (20)$$

must be valid. Without constant (20), the tangential mechanical stresses also vanish at infinity ($\sigma_{\theta\theta}(\infty) = \sigma_{\varphi\varphi}(\infty) = 0$), which seems to be reasonable from the physical point of view and can serve as an argument in favor of the validity of the solution obtained.

Thus, once boundary conditions (9) and (10) have been satisfied, we find

$$\sigma_{rr}(r) = \frac{1}{3}AD(1+\sigma)[I_1(\infty) - I_1(r)] - \frac{2AD}{3r^3}(1-2\sigma)I_2(r), \quad (21)$$

$$\sigma_{\theta\theta}(r) = \frac{1}{3}AD(1+\sigma)[I_1(\infty) - I_1(r)] + \frac{AD}{3r^3}(1-2\sigma)I_2(r). \quad (22)$$

Formulas (21) and (22) are derived in general form for an arbitrary density of forces $F(r)$ for the first time.

It can be shown that tensor $\sigma_{rr}(r)$ (21) may change sign as the coordinate of the point of observation is varied. In other words, both stretched and compressed regions are present in the focal volume. Tensor $\sigma_{\theta\theta}(r)$ (22) is however always positive; i.e., the tangential stresses always stretch the sample. This is in agreement with the original concept of the mechanism behind the *h*-layer breakdown [1, 3]. It seems likely that these tensile stresses are largely responsible for the damage of the crystal in the region of laser radiation focusing. Note that the stresses specified are unrelated to the lattice heating. They appear instantaneously immediately after the appearance of the hot electron cloud and adiabatically follow changes in the cloud parameters.

The method used for solving the elasticity equation casts some doubt on the existence of extreme points discovered in the early solution to the problem found with the mean-value theorem [1]. Taking into account

(22) and writing the extremum condition for the function $\sigma_{\theta\theta}(r)$ in the form

$$\frac{d\sigma_{\theta\theta}(r)}{dr} = 0, \tag{23}$$

we come to the relationship equivalent to (23):

$$I_2(r) + \frac{\sigma}{1-\sigma} r^4 F(r) = 0. \tag{24}$$

Equation (24) applies to any model of forces $F(r)$. However, all the components of this equation are nonnegative. This means that the only extreme point is the center of the focal spot. We have already mentioned that the study of this center is physically incorrect. Another "suspicious" point is the inflection of the function $\sigma_{\theta\theta}(r)$, as was mentioned earlier [3]. We will consider mechanical stresses near this point (here, the stresses are slightly different from the formal results obtained at $r = 0$). The extrema of the radial stresses $\sigma_{rr}(r)$ may be of some significance. However, these are most probably the points of minimum, rather than maximum, and so influence the total mechanical stress only slightly. The in-depth study of this problem is possible only after the model of forces $F(r)$ has been strictly defined.

DENSITY OF VOLUME FORCES INDUCED BY NONEQUILIBRIUM ELECTRONS

In [1], a model of density of forces $F(r)$ was touched upon to a minor extent. Here, we consider this point in detail. Two approaches to the description of these forces can be distinguished.

The first, most consistent approach implies the solution of nonequilibrium quantum kinetic equations including all effects significant for the problem. These are equations like those used in [4, 5] taking into account the interaction of electrons with intense light and scattering centers (phonons and impurities). They should be supplemented by similar equations for other (phonon and impurity) crystal subsystems. When combined, these kinetic equations must take into account the spatial inhomogeneity of all distribution functions. Then, the desired force $\mathbf{F}(\mathbf{r})$ can be found from the collision integral

$$\mathbf{J}_{\text{col}} = \frac{d}{dt}\langle \mathbf{p} \rangle. \tag{25}$$

The available apparatus of quantum statistics enables us to write expression (25) in explicit form. However, the mathematical solution of the problem that makes it possible to find a specific form of $\mathbf{F}(\mathbf{r})$ is still more complicated than the solution of the equations in [4, 5]. Therefore, another, less strict method to determine $\mathbf{F}(\mathbf{r})$ should be looked for.

The second approach assumes [1] that some quasi-steady-state distribution of hot electrons, $n_e = n_e(r)$, and some distribution of their "temperatures," $T_e = T_e(r)$, have already been found. Since Bloch electrons in a perfect crystal do not interact with the lattice, lattice imperfections alone should be taken into account (let them be residual impurities uniformly distributed over the crystal). When scattered by the imperfections, electrons transfer a certain momentum to the lattice. This momentum per unit volume and unit time equals the desired value of $\mathbf{F}(\mathbf{r})$.

Mathematically, the aforesaid takes the form

$$\mathbf{F}(\mathbf{r}) = \int \mathbf{p}_{\text{sc}} f(\mathbf{p}, \mathbf{r}) w(\mathbf{p}) d\mathbf{p}, \tag{26}$$

where $\mathbf{p}_{\text{sc}}$ is the momentum transferred to the lattice per one quantum-mechanical act of scattering and $w(\mathbf{p})$ is the probability of a nonequilibrium electron being scattered by a given lattice imperfection per unit time. The distribution function of nonequilibrium electrons, $f(\mathbf{p}, \mathbf{r})$, is assumed to be normalized to their density.

For our purposes, we only need to consider the spherically symmetric electron distribution and take into account that only one third of electrons moving in the radial direction transfers the momentum $\mathbf{p}_{\text{sc}}$ to the lattice. Then, instead of (26), we have

$$\mathbf{F}(\mathbf{r}) = \frac{1}{3}\frac{\mathbf{r}}{r}\int |\mathbf{p}| f(\mathbf{p}, \mathbf{r}) w(\mathbf{p}) d\mathbf{p}. \tag{27}$$

In view of the relation between the scattering probability and the total scattering cross section $\sigma_i N_i$ (where σ_i is the cross section of electron scattering by a lattice imperfection),

$$w(\mathbf{p}) = \frac{\sigma_i N_i}{V} v = \frac{\sigma_i N_i}{mV} p, \tag{28}$$

where N_i is the total number of scattering centers, V is the crystal volume, $N_i/V = n_i$ is the density of the centers, and m is the mass of an electron, we arrive at

$$\mathbf{F}(\mathbf{r}) = \frac{2}{3}\frac{\mathbf{r}}{r}\sigma_i n_i \int \frac{p^2}{2m} f(\mathbf{p}, \mathbf{r}) d\mathbf{p}. \tag{29}$$

Here, we took into consideration that the cross section σ_i depends on the electron momentum only slightly.

The integral in (29) is the average electron kinetic energy

$$\left\langle \frac{p^2}{2m} \right\rangle = \frac{3}{2} T_{\text{e}}(r) n_{\text{e}}(r), \tag{30}$$

where the "temperature" T_{e} has the dimension of energy.

Eventually, the formula for the radial component of the force density (other components are absent) takes the form

$$F(r) = \sigma_i n_i n_{\text{e}}(r) T_{\text{e}}(r). \tag{31}$$

Such an expression for the force density was used in [1], where it was derived somewhat differently, i.e.,

through the effective electron gas pressure on lattice imperfections with the scattering cross-section σ_i.

ESTIMATES AND CONCLUSIONS

To estimate whether the mechanical stresses $\sigma_{\theta\theta}$ can readily lead to the fracture of the sample subjected to a focused laser radiation, let us take the set of parameters used in [1] and consider the natural fracture criterion $\sigma_{\theta\theta} > \sigma_{max}$ (where σ_{max} is the tensile strength). As in [1], we assume that

$$n_e(r) = n_{e0}\exp\left\{-\frac{r^2}{a^2}\right\} \tag{32}$$

and

$$T_e(r) = T_{e0}\exp\left\{-\frac{r^2}{a^2}\right\}, \tag{33}$$

where n_{e0} and T_{e0} are the values typical of the center of the focal spot and a is the parameter of the Gaussian distribution of the laser radiation intensity over the focal spot.

Substituting expressions (32) and (33) into formula (22) and setting $b = a/\sqrt{2}$, we get

$$\sigma_{\theta\theta}(r) = \frac{Bb}{6(1-\sigma)}\left\{\sqrt{\pi}(1-\sigma)\left[1-\Phi\left(\frac{r}{b}\right)\right] + \frac{b^3}{r^3}(1-\sigma)\left[1-\left(1+\frac{r^2}{b^2}\right)\exp\left\{-\frac{r^2}{b^2}\right\}\right]\right\}. \tag{34}$$

In the framework of our model, result (34) has been obtained without any mathematical approximations and is exact in this sense.

Let us estimate tangential mechanical stress (34) by using the same set of parameters as before [1], i.e., by taking $\sigma_i = 10^{-14}$ cm^2, $n_i - 10^{15}$ cm^{-3}, $\sigma = 0.25$, $n_{e0} = 3 \times 10^{22}$ cm^{-3}, $T_{e0} = 0.5$ eV $= 8 \times 10^{-13}$ erg, and $a = 0.1$ cm. Then, at the inflection point, $r_0 = 0.8a$ and $\sigma_{\theta\theta} \approx 1.1 \times 10^9$ dyn/cm^2, which is roughly 60% higher than in [1]. With regard for the great difference between the methods of calculation, the agreement seems to be fairly good.

The value found exceeds typical tensile strengths. Therefore, the mechanism proposed may be responsible for the optical damage of pure transparent dielectrics.

Thus, we come to the following conclusions.

(1) The problem of elastic stresses in the spherically symmetric focal region of a pure transparent dielectric exposed to intense laser radiation is stated and exactly solved.

(2) The model for the density of forces acting on the crystal lattice due to a spatially nonuniform ensemble of nonequilibrium electrons laser-heated to high "temperatures" is refined and supplemented. These forces are by no means related to the lattice heating or thermal stresses in the lattice.

(3) The preliminary estimates [1], which show that the mechanism proposed does provide the optical damage of transparent dielectrics, are confirmed with a sufficiently high accuracy.

(4) The validity of using the natural criterion for optical damage, $\sigma_{\theta\theta} \geq \sigma_{max}$, is confirmed.

(5) The role of electron scattering by phonons, as well as on the sample surface, still remains unclear. The solution of the problem in the cylindrical coordinate system and a comparison of theoretical results with data derived with specially designed experiments where short-focus and long-focus lenses are used for radiation focusing seem to be of great importance.

(6) The mechanism considered is inertia-free. The volume forces $F(r)$ appear simultaneously with the electron cloud generation; therefore, this damage mechanism can be expected to be valid under the action of femtosecond laser pulses as well.

REFERENCES

1. V. N. Strekalov, Pis'ma Zh. Tekh. Fiz. **26** (24), 19 (2000) [Tech. Phys. Lett. **26**, 1081 (2000)].
2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity* (Nauka, Moscow, 1965; Pergamon, New York, 1986).
3. V. N. Strekalov, Cand. Sc. Dissertation (Moscow State Engineering Physics Institute, Moscow, (1975); Tr. MFTI, Ser. "Radiotekh. Élektron.", No. 9, 3 (1975).
4. É. M. Épsteĭn, Fiz. Tverd. Tela (Leningrad) **11**, 2732 (1969) [Sov. Phys. Solid State **11**, 2213 (1969)].
5. V. N. Strekalov, Fiz. Tverd. Tela (Leningrad) **15**, 1373 (1973) [Sov. Phys. Solid State **15**, 930 (1973)].

Translated by M. Lebedev

Technical Physics, Vol. 47, No. 9, 2002, pp. 1145–1151. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 80–86.

OPTICS, QUANTUM ELECTRONICS

Effect of Defects on the End Face of a Fiber on the Radiation Input Efficiency

D. V. Kiesewetter and V. I. Malyugin

St. Petersburg State Technical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia

e-mail: Kiesewetter@mailbox.alkor.ru

Received September 21, 2001

Abstract—The effect of optical nonuniformities on the end of a fiber on optical radiation input is analyzed. Simple relationships for angular characteristics and input efficiency are derived. The analytic expressions are compared with experimental data. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The quality of fiber end faces has an effect on the parameters of fiber joints and radiation propagation in the fiber. Defects on the ends may appear during the fabrication process and exploitation. The end faces may be damaged mechanically (scratches and surface roughness) or for other reasons, such as surface contamination; the presence of moisture on the fiber end faces, which condenses under drastic temperature drops; or the action of intense laser radiation.

A number of works [1–5] have been devoted to the effect of end face defects on the parameters of fiber joints. In this work, we generalize available theoretical and experimental data on the efficiency of input of a plane electromagnetic wave into a fiber with allowance for an optical nonuniformity of the end face.

The problem of calculating the radiation power in a fiber is divided into two parts: the determination of the light scattering indicatrix (SI) and the calculation of the fiber-filtered optical power. The first subproblem is associated with the effect of nonuniformities and contaminations present on the fiber optical surfaces on radiation scattering. It was attacked by numerically simulating and measuring the SI on plane-parallel optical plates made of the same material as the fiber core. On the test plates, researchers measured the surface roughness and determined the SI and attenuation of transmitted radiation using profilometry techniques. Results thus obtained were applied for the analysis of radiation input into fibers with different aperture angles and core diameters.

APPROXIMATION OF LIGHT SCATTERING INDICATRIX ON ROUGH SURFACES

We represent the scattering indicatrix $f(\gamma, \theta, \gamma', \theta')$ of a plane monochromatic electromagnetic wave on a rough surface normalized to the incident power F_0 as the sum of two, nonscattered and diffusely scattered, components [6, 7]:

$$f(\gamma, \theta, \gamma', \theta') = k_n\delta(\gamma - \gamma')\delta(\theta - \theta') + k_d f_d(\gamma, \theta, \gamma', \theta'),$$

where γ, θ, γ', and θ' are the azimuth and radial angles for the incident and scattered waves, respectively; $k_n = F_n/F_0$; $k_d = F_d/F_0$; and F_n and F_d are the powers of the nonscattered and diffusely scattered indicatrix components at a normal angle of incidence ($\gamma = 0$).

When radiation enters a fiber, the scattered radiation distribution in a narrow angular range ($\gamma < 15°$, according to the aperture angle of the fiber) is usually taken into account. Within this range, the diffuse component of the SI, f_d, can be approximated by the Gaussian function. Then, in a cylinder coordinate system,

$$f(\gamma, \theta, \gamma', \theta') = k_n\delta(\gamma - \gamma')\delta(\theta - \theta') + \frac{k_\sigma}{\pi\sigma^2}\exp\left[-\frac{1}{\sigma^2}(\gamma^2 - (\gamma')^2) - 2\cos(\theta - \theta')\gamma\gamma'\right], \quad (1)$$

where σ is the SI half-width, $k_\sigma = F_\sigma/F_0$, and F_σ is the power scattered in accordance with the Gaussian distribution.

The coefficients k_n, k_σ, and k_d are determined from SIs measured on plane-parallel plates with regard for the coefficient of Fresnel reflection at the interface and statistical characteristics of the scatterer surface relief. The use of the Gaussian approximation makes it possible to derive simple expressions for the efficiency of radiation input into light guides with damaged end faces and set typical relationships.

If the input efficiency (IE) is estimated for a wide range of angles of entry γ (for example, when radiation sources have a wide radiation pattern), an expression for the SI should include the shading effect [8], SI variation with angle of incidence [6], and an increase in k_n for $\gamma \longrightarrow \pi/2$. With such an approach, the calculations become very tedious, and, moreover, an adequate scat-

terer model is difficult to select. In this case, it is appropriate to use the relative transmission coefficient k_s, which is experimentally found as the ratio of the total radiation power passed through a plane-parallel plate with a scattering entrance surface, F_s, to the power passed through a nonscattering plate, F_0:

$$k_s = F_s/F_0. \quad (2)$$

For phase scattering, the difference between k_s and unity is due to the reflection of the scattered radiation incident on the second (nonscattering) surface of the plate at an angle larger than the angle of total internal reflection γ_i. If the end face of a fiber serves as a diffuser, the radiation scattered through angles larger than γ_i is not captured by core and cladding modes. In this situation, the value of k_s can be viewed as the maximal achievable input efficiency for a light guide with the scattering end face. It will be shown below that the introduction of the coefficient k_s found experimentally allows one to take into account the non-Gaussian shape of the actual SI and, accordingly, to improve the IE estimate.

RADIATION POWER FILTERED BY A FIBER

Considering a fiber as a filter with the transfer characteristic $G(\gamma, \theta)$, we can write the expression for its output optical power $P(\gamma, \theta)$ in the form

$$P(\gamma, \theta) = N \int_0^{\pi/2} \int_0^{2\pi} G(\tilde{\gamma}'\tilde{\theta}')\, f(\gamma, \theta, \tilde{\gamma}', \tilde{\theta}')\tilde{\gamma}' d\tilde{\gamma}' d\tilde{\theta}', \quad (3)$$

where N is a normalizing factor and $\tilde{\gamma}'$ and $\tilde{\theta}'$ are the scattering angles "inside" the fiber in a medium with a refractive index n_1, which equals that of the core at the point of scattering.

The scattering parameters and angular characteristics [2, 9] are determined for the angles γ' and θ' measured in free space; therefore, it is appropriate to perform integration in (3) in the same angular range using the conversion relationship $\sin\gamma = n_1 \sin\tilde{\gamma}'$. Hereafter, by the term "variation of input efficiency" we mean the relative variation of the output power (transferred by both core and cladding modes) under different input conditions. Any variation of the radiation power because of different dampings of various modes will also be considered as the IE variation.

Integration in (3) implies that the input power is additive with respect to angle of incidence and wave scattering. Such an approach, which is beyond question if the source is noncoherent, is admissible for coherent waves if their interference provides a statistically random speckle pattern with a characteristic speckle size $S_s \ll \pi a^2 \leq S_b$, where a is the fiber core radius and S_b is the surface area of a beam illuminating the end face. In determining the transfer characteristics, it is also assumed that the differential mode damping, scattering, and mode coupling are linear. If the radiation is introduced with a focused beam, the transfer characteristic (or the angular input characteristic (AIC) in terms of relative quantities) should be measured under similar conditions. The model used also enables the mode composition of the radiation to be determined if the transfer function is expressed through the excitation coefficients for a given waveguide mode or group of modes [1].

Our approach also applies to the estimation of the scattered power filtered by various optoelectron devices even if the source has a complex radiation pattern. However, this problem calls for special consideration.

The input power can be split into two parts associated with the nonscattered (directed), P_n, and diffuse, P_d, components of the SI. Accordingly, the variation of the input characteristics due to scattering by the entrance end face also depends on two components; however, the variation of the angular relationships is a function of P_d alone. Using the relationships for Bessel functions [10]

$$\frac{1}{2\pi}\int_0^{2\pi} \exp\left(\cos(\theta - \theta')\frac{2\gamma\gamma'}{\sigma^2}\right)\exp(i\theta 0)d\theta' = I_0\left(\frac{2\gamma\gamma'}{\sigma^2}\right),$$

$$\int_0^{\infty} I_0\left(\frac{2\gamma\gamma'}{\sigma^2}\right)\exp\left[-\frac{1}{\sigma^2}(\gamma^2 - (\gamma')^2)\right]\gamma' d\gamma = \pi\sigma^2,$$

we rearrange expression (3) into the form

$$\frac{P_{out}}{P_0}N_t = k_n g(\gamma_r) + \frac{2k_\sigma}{\sigma_r^2}\int_0^{\infty} g(\gamma_r')I_0\left(\frac{2\gamma_r\gamma_r'}{\sigma_r^2}\right) \times \exp\left(-\frac{\gamma_r^2 - (\gamma_r')^2}{\sigma_r^2}\right)\gamma_r' d\gamma_r', \quad (4)$$

where P_{out} is the output; P_0 is the input (the power incident on the end face); I_0 is the modified Bessel function; $\gamma_r = \gamma/\gamma_f$ and $\sigma_r = \sigma/\gamma_f$ are the relative angle of entry and the SI half-width normalized to the aperture angle of a fiber γ_f, respectively; $g(\gamma) = G(\gamma)/G(\gamma = 0)$ is the input angular characteristic; and N_t is a normalized factor dependent on the transfer properties of the fiber and radiator parameters.

Passing to the relative values of the power, for example, relative to the power for the nonscattering end face $p(\gamma) = P(\sigma, \gamma)/P(\sigma \longrightarrow 0, \gamma)$, or to the dependence of the power on the angle of entry, $p(\gamma) = P(\gamma)/P$ ($\gamma = 0$), one can analyze resulting effects without calculating the normalizing factor N_p.

Let us consider a plane wave entering a fiber with a diffusely scattering end face ($k_n = 0$ and $k_d = 1$) and estimate the IE for various AIC types. In particular, we

approximate the AIC by a step function,

$$g_{st}(\gamma_r) = \begin{cases} 1 \ldots \gamma_r \le 1 \\ 0 \ldots \gamma_r > 1, \end{cases}$$

and by a Gaussian function,

$$g_G(\gamma_r) = \exp(-\gamma_r^2).$$

For a wave propagating along the fiber axis ($\gamma = 0$), the IE

$$p_\sigma(\sigma_r, \gamma_r) = p(\sigma_r, \gamma_r = 0)/p(\sigma_r \longrightarrow 0, \gamma_r = 0)$$

is described by the simple analytic expressions

$$p_G(\sigma_r, \gamma_r = 0) = 1 - \exp\left(-\frac{1}{\sigma_r^2}\right) \quad \text{for} \quad g_{st}, \qquad (5)$$

$$p_\sigma(\sigma_r, \gamma_r = 0) = \frac{1}{1 + \sigma_r^2} \quad \text{for} \quad g_G. \qquad (6)$$

For the inclined incidence of the wave ($\gamma \neq 0$), the IE variation is calculated by numerically integrating (4) (Fig. 1). For high values of $(\gamma_r \gamma_r'/\sigma_r^2)$, the asymptotic formula

$$I_0\left(\frac{2\gamma_r\gamma_r'}{\sigma_r^2}\right)\exp\left[-\frac{1}{\sigma_r^2}\{\gamma_r^2 - (\gamma_r')^2\}\right]$$

$$\approx \exp\left[-\frac{1}{\sigma_r^2}\{\gamma_r^2 - (\gamma_r')^2\}\right]\left\{\frac{\sigma_r}{2\sqrt{\pi\gamma_r\gamma_r'}}\left(1 + \frac{\sigma_r^2}{16\gamma_r\gamma_r'}\right)\right\}$$

is used. The dependence

$$p_{\gamma,\sigma}(\sigma_r, \gamma_r) = p(\sigma_r, \gamma_r)/p(\sigma_r \longrightarrow 0, \gamma_r = 0)$$

virtually characterizes a change in the IE compared with the case of a nonscattering end face, and

$$p_\gamma(\sigma_r, \gamma_r) = p(\sigma_r, \gamma_r)/p(\sigma_r, \gamma_r = 0)$$

describes the IE variation with angle γ_r, that is, the AIC for a scattering end face. For the Gaussian AIC, the dependence $p_\gamma(\sigma_r, \gamma)$ can be obtained in the analytical form

$$p_\gamma(\sigma_r, \gamma_r) = \exp(-\gamma_r^2/(1 + \sigma_r^2)). \qquad (7)$$

Thus, when light is scattered by the entrance end face (Fig. 1), waveguide modes are excited in a wider range of angles of incidence of the plane electromagnetic wave. This is equivalent to an increase in the input aperture angle. Such an effect can be explained in simple physical terms. Upon diffuse scattering by the entrance end face, the angular aperture collects the scattered radiation if angles of entry $\gamma > \gamma_f$. The AICs for a fiber with a scattering end are similar to those for a fiber with a nonscattering end if the aperture angle of the latter, γ_f^*, is larger but its IE lower. Figure 2a demonstrates the dependences of the equivalent aperture angle of entry $\chi_e = \gamma_f^*/\gamma_f$ on the SI half-width that are obtained by solving the equation $p_\gamma(\sigma_r, \gamma_r) = u$ (u is the aperture level, $0 < u < 1$). If $\sigma_r \gg 1$, we have the simple asymptotic solution $\chi_e = \sigma_r(-\ln u)^{1/2}$, which applies even to $\sigma_r > 1.5$. For $\sigma_r \gg 1$, the dependence $\chi_e(\sigma_r)$ is a monotonically increasing function, while for $\sigma_r < 1$, it depends on u.

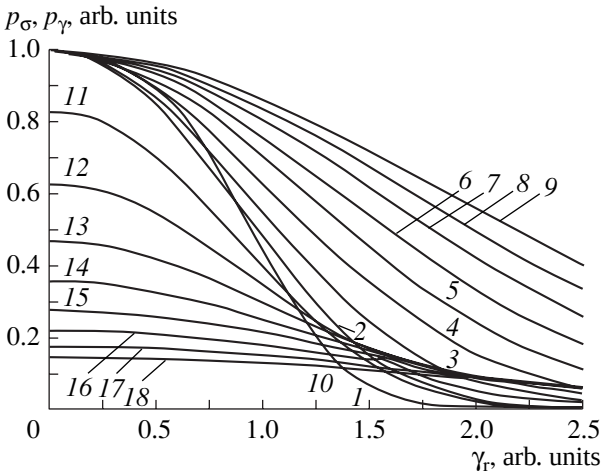


Fig. 1. Variation of the (*1–10*) AICs $p_\gamma(\sigma_r, \gamma_r) = p(\sigma_r, \gamma_e)/p(\sigma_r, \gamma_r = 0)$ and (*11–18*) input efficiency $p_\sigma(\sigma_r, \gamma_r) = p(\sigma_r, \gamma_r = 0)/p(\sigma_r \longrightarrow 0, \gamma_r = 0)$ in comparison with those for the nonscattering entrance end face versus relative angle of entry $\gamma_r = \gamma_f$ for a fiber with the step AIC. $\sigma_r = 0.25$ (*1*, *11*), 0.5 (*2*, *12*), 1 (*3*, *13*), …, and 2.5 (*10*, *20*). Curves *1*, *11* and *2*, *12* coincide.

To estimate the effect of the AIC shape on $\chi(u)$, the equation $p_\gamma(\sigma_r, \gamma_r) = u$ was solved for step, Gaussian, triangular, and trapezoidal AICs (Fig. 2b). For the Gaussian AIC, the solution has the analytical form

$$\chi = (\gamma_f^2 + \sigma^2)^{1/2}, \qquad (8)$$

where χ, γ_f, and σ are defined by the $1/e$ level.

For any of the AIC shapes (Fig. 2b), the values of $\chi(\sigma)$ differ from those obtained by (8) by 10–20%. Thus, relationship (8) is fairly accurate for virtually all real types of optical waveguides if $P_d \gg P_n$. Note that $\chi_e(\sigma_r)$ at $\sigma_r \longrightarrow 0$ depends on u for all of the AICs except for the step function. As the SI widens, the relationship $p_\gamma(\sigma_r, \gamma_r)$ depends to an ever decreasing extent on the fiber AIC but more and more strongly on the scatterer parameters. However, the possibility of $\chi(\sigma)$ being approximated by function (8) for various light guides with a diffusely scattering end face implies that this expression is also applicable in the more general case $k_n \neq 0$.

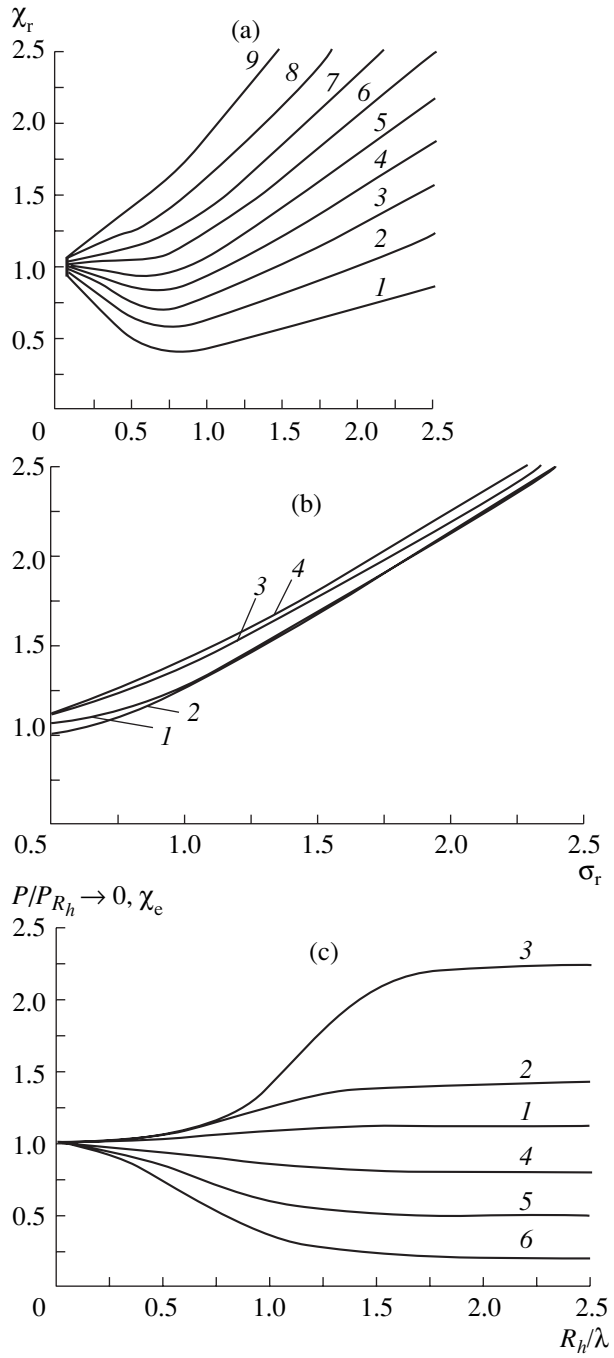


Fig. 2. (a) Relative equivalent aperture angle χ_r vs. the relative half-width σ_r of the radiation pattern for the aperture level u = 0.9 (*1*), 0.8 (*2*), 0.7 (*3*), …, 0.1 (*9*). (b) Relative equivalent aperture angle χ_e vs. the relative half-width σ_r of the radiation pattern for fibers with (*1*) step, (*2*) trapezoidal, (*3*) triangular, and (*4*) Gaussian AICs. Curve *4* coincides with the approximation $\chi_e = (1 + \sigma_r^2)^{1/2}$. (c) (*4–6*) The relative input efficiency $P/P(T_h \longrightarrow 0)$ and (*1–3*) equivalent aperture angle χ_e vs. relative surface roughness height R_h/λ: $\sigma_r = \gamma_f/2$ (*1*, *4*) γ_f, and $2\gamma_f$ (*3*, *6*).

THE EFFECT OF END FACE ROUGHNESS ON THE INPUT EFFICIENCY

Let us consider the variation of the input characteristics with the rms height R_h of surface roughness with regard for the coefficients k_n and k_σ. For light guides with relatively smooth surfaces ($\gamma_f \ll \sigma$ and $P_n \gg P_d$), the IE depends on the nonscattered component of the indicatrix, k_n. As the roughness increases, the diffuse component of the indicatrix becomes significant and both components should be taken into account in (4). If $R_h \gg \lambda/2$, the second term $P_n \ll P_d$, related to scattering by the entrance end face, dominates. Using k_n, k_d, and σ measured on plane-parallel plates, one can calculate the angular characteristics and efficiency variation as functions of R_h [4, 11]. Of interest is the theoretical dependence of the relative efficiency $p_r(\gamma, R_h) = p(\gamma, R_h)/p(\gamma, R_h \longrightarrow 0)$ for various degrees of roughness. Let us find the R_h dependence of $p_r(\gamma, R_h) = p(\gamma, R_h)/p(\gamma, R_h \longrightarrow 0)$ using the expression [6, 8]

$$k_n(R_h) = \exp\left(-\frac{\mu R_h^2}{\lambda^2}\cos^2\gamma\right) \quad (9)$$

as an estimate of the nonscattered component of the indicatrix. Here, λ is the wavelength, $\cos^2\gamma \approx 1$, and $\mu \approx \pi^2$ (for the reflected wave, $\mu^{(r)} \approx 16\pi^2$).

If the AIC shape is Gaussian, formulas (4) and (9) yield the expression for the AIC when the end face is rough,

$$p_\gamma(\sigma_\gamma, \gamma_\gamma) = k_n^* \exp(-\gamma_r^2) + \frac{k_\sigma^*}{1 + \sigma_r^2}\exp\left(-\frac{\gamma_r^2}{1 + \sigma_r^2}\right),$$

and the expression for the IE variation relative to the IE for the smooth (nonscattering) end face,

$$p/p(\sigma \longrightarrow 0) = k_n^{**} + \frac{k_\sigma^{**}}{1 + \sigma_r^2}\exp\left(\frac{\sigma_r^2\gamma_r^2}{1 + \sigma_r^2}\right),$$

where $k_n^* = k_n/(k_n + k_\sigma/(1 + \sigma_r^2))$, $k_\sigma^* = k_\sigma/(k_n + k_\sigma/(1 + \sigma_r^2))$, $k_n^{**} = k_n/(k_n + k_\sigma)$, and $k_\sigma^{**} = k_\sigma/(k_n + k_\sigma)$.

It should be noted that, when $\gamma_r > 1$, the relative IE $p/p(\sigma \longrightarrow 0)$ may exceed unity (that is, the IE may exceed that for a nonscattering entrance end face). In the case of normal incidence ($\gamma = 0$), we have

$$p/p(\sigma \longrightarrow 0, \gamma = 0) = k_n^{**} + k_\sigma^{**}/(1 + \sigma_r^2).$$

In the simplest case $k_\sigma = 1 - k_n$, $k_n^{**} = k_n$, and $k_\sigma^{**} = k_\sigma$. The equivalent aperture angle χ_e can be calculated only by numerically solving the equation $p(\gamma, R_h) = u$. The dependences $p_k(R_h)$ and $\chi_e(R_h)$ for various ratios between the SI half-width and aperture angle ($\gamma_f = \sigma_f/2$, $\gamma_f = \delta_f$, and $\gamma_f = 2\sigma_f$) are demonstrated in Fig. 2c. The

dependence $\chi(R_h)$ can be subdivided into three stages. At the initial stage ($R_h/\lambda < 0.1$), the scattering is relatively low and does not markedly affect the value of the equivalent aperture angle ($\chi_2 \approx 1$) and the IE. At stage 2 ($0.1 < R_h/\lambda < 0.5$), the curves $\chi(R_h)$ and $p(R_h)$ vary most significantly. At the third stage ($R_h > 0.5$), the surface scatters the light diffusely ($k_n \approx 0$); therefore, as R_h increases further, the aperture angle and the IE remain unchanged. The asymptotic values of $p_k(R_h)$ and $\chi_e(R_h)$ at $R_h/\lambda > 0.5$ are given by expressions (6) and (8).

If $R_h/R_c \approx$ const, where R_c is the correlation distance of surface roughness, the SI half-width σ can be considered as R_h-independent, $\sigma(h)$ = const, and, as a consequence, the function $\chi(R_h)$ is considered to vary slightly with the microrelief height. For actual rough surfaces obtained by grinding with abrasives, both R_h and the SI half-width grow with increasing grain size because of a change in the correlation distance of surface roughness [5]. Accordingly, at $R_h/\lambda > 0.5$, the function $\chi(R_h)$ will slightly increase, while $p_k(R_h)$ decrease, as was observed in [2].

EXPERIMENTAL STUDY OF INPUT EFFICIENCY

We studied the input characteristics of fiber light guides with a rough end face, R_h = 0.1–0.8 μm. The end faces of the fibers and reference plane-parallel plates were ground by powders with different grain sizes. Grinding conditions, roughness statistical characteristics, and SI parameters can be found elsewhere [4, 11–13]. Using the angular characteristic measurement technique [2, 9], we found the dependences $\chi_e(R_h)$ and $p(R_h)$ for various types of fiber light guides at wavelength λ = 0.63 μm (Figs. 3, 4) by using analytical dependence (9) in contrast to [1, 3], where the experimental values of k_n and k_σ were used. The dependence $\sigma(R_h)$ was defined as a linear function $\sigma(R_h) = \alpha + \beta R_h$ with the coefficients α and β obtained by approximating the experimental data in [4, 11]: $\alpha \approx 3.7$ and $\beta \approx 9.05$ (σ is given in degrees and R_h, in micrometers). The analytical and experimental data for $\chi(R_h)$ are found to be in good agreement. For $p(R_h)$, our simple model gives an overestimated value. For light guides *1–3* in Fig. 3, the rms deviations for $p(R_h)$ and $\chi(R_h)$ are 19, 10, and 17% and 5–6%, respectively. The discrepancies are explained largely by the use of the approximation $k_\sigma = 1 - k_n$, since for real SIs, $(k_\sigma + k_n) < 1$, as indicated by the measurement of the relative transmission coefficient k_s.

The coefficient k_s was measured on the reference plane-parallel plates with a rough entrance face. The radiation transmitted through the plate was detected by a photoelectric multiplier with a high surface area of the photocathode. The power transmitted through the rough plate was lower than that passed through the plate with the smooth surface (curve *7* in Fig. 4) because of the reflection from the second face of the rough plate at angles of incidence larger than γ_i. Estimates made with the Beckmann formula [4, 6, 7] indicate that the power scattered through angles $\gamma' > \gamma_i$ must be three to five orders of magnitude lower than the value found experimentally. This is likely to be associated with the non-Gaussian statistics of glass surface roughness and the effect of subsurface scattering. The formula for IE can be refined by putting $k_\sigma = k_s(1 - k_n)$;

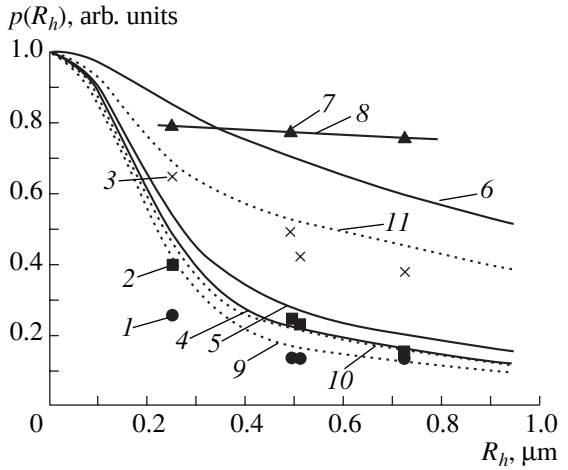


Fig. 3. Input efficiency as a function of the rms surface relief height R_h for various fibers: (*1–3*) experiment, (*4–6*) calculation by (6), (*7*) k_s found experimentally, (*8*) linear approximation of k_s, and (*9–11*) calculation by (10). Curves *1*, *4*, and *9* correspond to a fiber with a step profile of the refractive index for a core radius a = 100 μm and an aperture angle relative to the 1/*e* level $\gamma_f = 4.34°$; *2*, *5*, and *10*, to a 50/125 gradient fiber with a = 25 μm and $\gamma_f = 5°$; and *3*, *6*, and *11*, to "polymeric" quartz with a = 100 μm and $\gamma_f = 12.4°$.

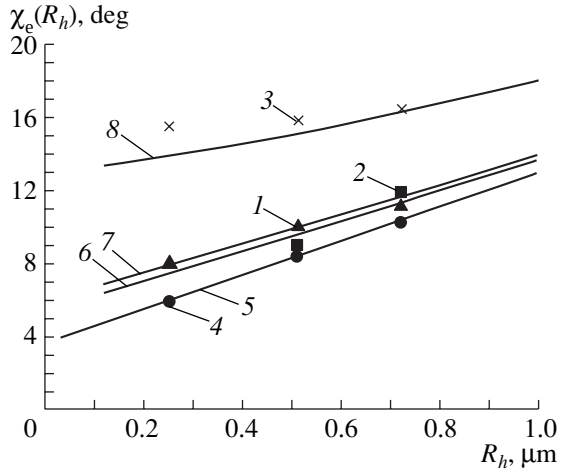


Fig. 4. Equivalent aperture angle χ_e vs. rms surface relief height R_h: (*1–3*) experiment; (*4*) SI half-width [11]; (*5*) approximation of data points *4*; and (*6–8*) calculation by (8). $\gamma_f = 4.34°$ (*1*, *6*), 5° (*2*, *7*), and 12.4° (*3*, *8*).

then,

$$p(R_h) = k_n(R_n) + \frac{[1 - k_n(R_n)]k_s}{1 - \sigma^2/\gamma_f^2}. \quad (10)$$

Taking $k_s = 0.75$ for all roughness types and estimating $p(R_h)$ with (10), we come to a better coincidence with the experiment (curves *9–11*). For $p(R_h)$, the discrepancies now are 13, 6, and 7%, respectively.

Having found experimentally the specular component $k_n^{(r)}$ of the SI for the radiation reflected from the fiber end face, one can estimate k_n for the radiation transmitted. Since $\mu \approx \pi^2$ and $\mu^{(r)} \approx 16\pi^2$, we have

$$k_n = (k_n^{(r)})^{1/16}. \quad (11)$$

For example, if the coefficient $k_n^{(r)}$ of specular reflection from a uniformly rough end face exceeds 0.44, then $k_n > 0.95$; that is, scattering losses are no more than 5%.

It is of interest to estimate a change in the power introduced into the fiber for the case of amplitude–phase scattering, for example, if the entrance face is partially contaminated. For a weakly contaminated surface (assuming that contaminants are randomly distributed over the end face; that is, $S_{con} \ll S_{core}$, where S_{con} is the surface area of contaminated regions and $S_{core} = \pi a^2$ is the core cross section surface area, and neglecting their scattering properties), we can write

$$k^{(a)} = S_{corl} - S_{con}/S_{corl}, \quad k_i^{(s)} = k^{(a)}k_i \quad i = \{n, \sigma, d, s\}, \quad (12)$$

where $k^{(a)}$ is the amplitude transmission coefficient; $k_n^{(s)}$, $k_\sigma^{(s)}$, and $k_d^{(s)}$ are the resulting coefficients of directed transmission and diffuse scattering; and $k_s^{(s)}$ is the resulting relative transmission coefficient.

The power introduced into a fiber can also be estimated with relationship (4) or (10) if $k_n^{(s)}$ and $k_\sigma^{(s)}$ are substituted for k_n and k_σ and inhomogeneities are uniformly distributed over the surface. However, the latter condition for rough surfaces frequently fails (contaminants are concentrated largely on recessed regions of the surface). If contamination increases the skewness and kurtosis of the surface profile height distribution, the scattering properties of such a surface are insignificant and relationship (6) may underestimate $p(R_h)$ if the concentration of inhomogeneities is low [4]. It should be noted that relationship (11) cannot be used for amplitude–phase scattering and statistically inhomogeneous surfaces, for example, surfaces with scarce defects.

In the case of a wavy surface, the SI can be calculated by applying the geometrical model of refraction of light by such surfaces and setting the coefficient k_s to be equal to unity.

Let us make estimates for other types of end face defects, for example, for a skewed surface. At small skew angles ε, the SI of such an end face ($k_n = 1$, $k_d = 0$) can be represented as

$$f(\gamma, \theta, \gamma', \theta') = \delta(\gamma - \gamma' + \varepsilon)\delta(\theta - \theta'),$$

hence,

$$\frac{P(\varepsilon)}{P(0)} = \frac{G(\varepsilon)}{G(0)} \quad \text{or} \quad p(\varepsilon) = g(\varepsilon),$$

where $G(0)$, $P(0)$ and $G(\varepsilon)$, $P(\varepsilon)$ are the angular characteristics and the input powers for normal ($\varepsilon = 0$) and skewed end faces, respectively.

CONCLUSION

We showed that scattering by the end face of a multimode fiber may both decrease the input efficiency and change the input angular characteristics. Irrespective of the type of the fiber and its transfer characteristics, the effect of end face scattering can be estimated by using the Gaussian approximation for the diffuse component of the SI and the delta approximation for its nonscattered (directed) component. This results in simple analytical expressions. If the theoretical values of the coefficient k_n are used (which imply the Gaussian model of rough surface), it is necessary to introduce a correction coefficient k_s that allows for the non-Gaussian shape of the actual SI. For abrasive-ground quartz glass with rms surface relief heights in the range 0.2–0.8 μm, k_s is close to 0.75; accordingly, the actual input efficiency equals 75% of the theoretical value.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (grant no. 00-02-16903).

REFERENCES

1. D. V. Kizevetter and V. I. Malyugin, Zh. Tekh. Fiz. **56**, 207 (1986) [Sov. Phys. Tech. Phys. **31**, 124 (1986)].
2. D. V. Kizevetter and V. I. Malyugin, Opt. Spektrosk. **64**, 1139 (1988) [Opt. Spectrosc. **64**, 677 (1988)].
3. D. V. Kizevetter and V. I. Malyugin, Tr. Leningr. Politekh. Inst. im. M. I. Kalinina, No. 429, 91 (1989).
4. D. V. Kizevetter, M. Ya. Litvak, and V. I. Malyugin, *Problems of Physical Electronics-91* (Fiz.-Tekh. Inst. im. A. F. Ioffe, Leningr. Gos. Tekh. Univ., Leningrad, 1991), pp. 120–165.
5. D. V. Kizevetter, Candidate's Dissertation (Leningrad, 1989).
6. A. S. Toporets, *Optics of Rough Surface* (Mashinostroenie, Leningrad, 1988).

7. P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surface* (Pergamon, Oxford, 1963).

8. F. G. Bass and I. M. Fuchs, *Wave Scattering from Statistically Rough Surfaces* (Nauka, Moscow, 1972; Pergamon, Oxford, 1978).

9. D. V. Kizevetter and V. I. Malyugin, Opt.-Mekh. Promst. **56** (9), 48 (1989) [Sov. J. Opt. Technol. **56**, 577 (1989)].

10. G. N. Watson, *Treatise on the Theory of Bessel Functions* (Cambridge Univ. Press, Cambridge, 1945; Inostrannaya Literatura, Moscow, 1949).

11. D. V. Kizevetter and V. I. Malyugin, Opt.-Mekh. Promst. **54** (2), 13 (1987) [Sov. J. Opt. Technol. **54**, 77 (1987)].

12. D. V. Kizevetter and V. I. Malyugin, Opt.-Mekh. Promst. **56** (6), 33 (1989) [Sov. J. Opt. Technol. **56**, 362 (1989)].

13. D. V. Kizevetter, M. Ya. Litvak, and V. I. Malyugin, *Physics and Diagnostics of Components and Active Media of Electronics* (Leningr. Gos. Tekh. Univ., Leningrad, 1991), Tr. Leningr. Gos. Tekh. Univ., No. 436, pp. 87–90.

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1152–1157. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 87–92.

OPTICS, QUANTUM ELECTRONICS

1-THz Low-Noise SIS Mixer with a Double-Dipole Antenna

S. V. Shitov*, A. V. Markov*, B. D. Jackson, A. M. Baryshev**, N. N. Iosad***, J.-R. Gao***, and T. M. Klapwijk*****

** Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Mokhovaya ul. 11, Moscow, 101999 Russia*

e-mail: Sergey@hitech.cplire.ru

*** National Institute for Space Research of the Netherlands (SRON), Groningen, the Netherlands*

**** Delft University of Technology (DIMES), Delft, the Netherlands*

Received January 30, 2002

Abstract—A quasi-optical mixer containing two Nb/Al/AlO_x/Nb superconducting tunnel junctions integrated into a NbTiN/SiO_2/Al microstrip line is studied experimentally in the 800–1000 GHz frequency range. The mixer is developed as an optional front end of the heterodyne receiver operating in frequency band 3 or 4 and incorporated into the HIFI module of the Herschel space-borne telescope. The double-dipole antenna of the mixer is made of NbTiN and Al films; the quarter-wavelength reflector, of a Nb film. The mixer is optimized for the IF band of 4–8 GHz. The double-sideband noise temperature T_{RX} measured at 935 GHz is 250 K at a mixer temperature of 2 K and an IF of 1.5 GHz. Within 850–1000 GHz, T_{RX} remains below 350 K. The antenna pattern is symmetrical with a sidelobe level below –16 dB. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Low-noise heterodyne receivers of the terahertz range are necessary for designing high-efficiency space-borne telescopes, which are being developed for submillimeter spectral astronomy, for example, for the HIFI [1]. Upon tackling the problem of designing a high-sensitivity 1-THz receiver, it is reasonable to analyze the key parameters of superconductor–insulator–superconductor (SIS) mixers built on tunnel junctions.

Nb/Al/AlO_x/Nb tunnel-junction mixers are known as low-noise heterodyne converters. Their noise temperature on the order of hf/k_B (where h, f, and k_B are the Planck constant, frequency, and Boltzmann constant, respectively) [2] is limited by quantum fluctuations. Such mixers have been studied experimentally at 30–1500 GHz, and their noise temperature has been estimated at about $(2–3)hf/k_B$ at frequencies below 680 GHz (the niobium gap frequency [3, 4]). Above this value, the sensitivity decreases rapidly primarily because of increased losses in niobium tuning circuits. Theoretically, the frequency range of a niobium-based SIS heterodyne converter is limited by the doubled niobium gap frequency, i.e., approximately by 1300 GHz [5]. This is because the conversion efficiency is associated with the nonlinear tunnel current of quasi-particles in SIS junction and thus does not depend directly on losses in the feeding circuits.

Designing a wide-band SIS mixer that would cover a substantial part of the subterahertz range is a challenge, since the capacitance per unit area C of an SIS junction is high. Accordingly, the loaded Q of the circuit is also high (typically about 10 at 1 THz). The high value of Q poses at least two problems: the bandwidth narrows and the losses increase. The real part of the microwave impedance of a superconducting film [6] raises the loss in proportion to the current density squared, i.e., to Q^2. Theoretically, the system's Q factor can be lowered by reducing the resistivity of the tunnel barrier, i.e., by increasing J_c. Unfortunately, this is difficult to do in practice, because the I–V characteristic of the SIS junction degrades when $J_c > 10–15$ kA/cm^2 [7]. Another possibility of decreasing the microwave loss is to reduce the SIS junction area, which also lowers the current density in the feeding circuits. Thus, it is clear that, in general, quantum-sensitive SIS mixers are difficult to design for the frequency range where the loss in the feeding circuits is high. That is why, along with efforts in minimizing the high-frequency loss by increasing the tunnel current density and shrinking SIS junctions, advanced designs of low-loss tuning circuits based on novel materials [8–12] are of great importance for developing a terahertz-range quantum-sensitive mixer.

The capacitance of an SIS junction together with that of the integrated tuning circuits may hamper the matching of the mixer's output at the IF due to the high dynamic resistance of the junction. Under operating conditions, this resistance may be within 0.2–1.0 kΩ and even become negative [2], which leads to a high Q of the mixer's output, especially at high IFs. The problem of wide-band matching the mixer's output can be solved by decreasing the capacitance of all tuning and connecting circuits and applying a special matching transformer.

1063-7842/02/4709-1152$22.00

A high-frequency signal can be fed to the SIS mixer by two basic ways: through a rectangular waveguide or with the help of a quasi-optical antenna integrated with the SIS junction. The problems mentioned above are common to both designs. Specific problems may arise in fabricating precision mechanical parts of waveguide mixers. For example, the cross section of a single-mode terahertz waveguide is as small as 120×240 μm, which means that other parts such as a scalar horn, tuning pistons, a chip mounting channel, etc. must be fabricated with a micrometer precision. Auxiliary stages of chip preparation (especially dicing and grinding), as well as the assembling of a waveguide mixer, are also critical operations.

Unlike waveguide mixers, quasi-optical mixers can be fabricated on relatively large and easy-to-handle substrates, which are a part of the optical system. The parameters of these mixers depend primarily on the photolithography resolution, which is several fractions of a micron. It should be noted that the pattern of the antenna–lens system is defined by both the quality of the microwave lens and the accuracy of placing the antenna on the optical axis of the lens [13, 14]. The theoretical level of the first sidelobe of integrated lens antennas, which are most widely used, is about –18 dB, that is, slightly higher (worse) than that of scalar horns used with waveguide mixers. However, this value is adequate for most applications.

Double-slot and double-dipole antennas are widely used in planar receivers [10, 15–18]. When incorporated into an integrated lens antenna, these have similar patterns and close values of impedance. An advantage of a slot antenna is its wide ground plane, which can accommodate comparatively complex circuits near the antenna. At the same time, a double-dipole antenna was successfully used in a complex superconducting integrated receiver [16]. The back lobe of a slot antenna cannot be used as in the case of a dipole antenna, because the convergent reflector becomes inefficient. A double-dipole antenna is usually free of capacitive coupling elements, so that the average capacitance of the structure can be made lower, which is important when the IF is high.

The primary goal of this study was to develop and experimentally demonstrate a quasi-optical SIS mixer with an integrated lens antenna that meets the HIFI requirements for frequency bands 3 and 4 [1].

GENERAL APPROACH

The structure of choice is a twin-type mixer, which contains two Nb/Al/AlO_x/Nb SIS junctions connected in antiphase and a double-dipole antenna as shown in Fig. 1. This resonant structure of SIS junctions has been theoretically and experimentally shown to have a wider matching band and a lower loss than a circuit where the junction terminates a tuning microstrip line [10, 18]. The lower losses can be explained by the low density of

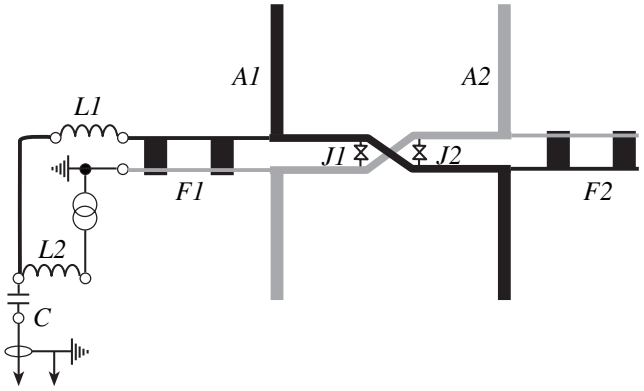


Fig. 1. Circuit diagram of a planar quasi-optical SIS mixer with a double-dipole (*A*1 and *A*2) antenna. The mixer comprises two SIS junctions (*J*1 and *J*2) that operate in antiphase and are parallel-connected in terms of IF current but series-connected in terms of microwave current. Inductance $L1 = 0.7$ nH is a tuning element at the IF output, *C* and *L*2 constitute the filter of the bias source, and *F*1 and *F*2 are the band-stop filters of the antenna.

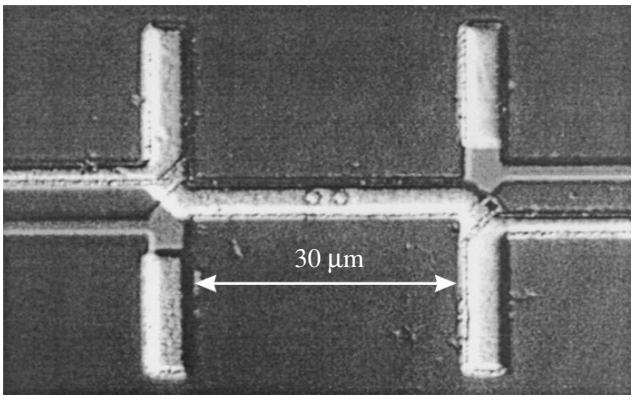


Fig. 2. SIS mixer with a double-dipole antenna on a silicon substrate. Two micron-size SIS junctions are seen at the center. Bright metal strips are aluminum; dark lines, NbTiN. The SiO_2 insulator between the electrodes is almost transparent.

the microwave current at the input of the impedance transformer, which is typical of twin SIS structures. To minimize the ohmic loss in the tuning circuit, we used microstrip lines made of NbTiN alloy, for which the superconductor critical temperature is about 16 K and the gap frequency is about 1 THz [19, 20]. Theoretically, such values of the parameters can provide very low losses at this frequency. The trade-off in our design is the microstrip line on the NbTiN/SiO_2/Al structure. First, the upper aluminum electrode minimizes the effect of thermal trapping and, as a consequence, eliminates the overheating of the junction, which is observed in all-NbTiN tuning circuits [19]. Second, SiO_2 is far from being the optimal underlayer for the regular growth of NbTiN films, which may cause too

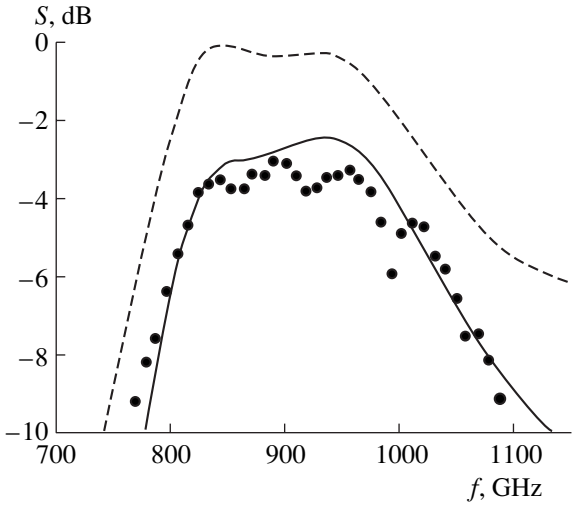


Fig. 3. Signal on the SIS junctions (circles) measured with a Fourier spectrometer and the computed values (*S*) (solid line). The signal level at the antenna (dashed line) is seen to be significantly higher than the signal applied to the SIS junctions because of ohmic losses in the aluminum conductors of the NbTiN/SiO_2/Al microstrip line.

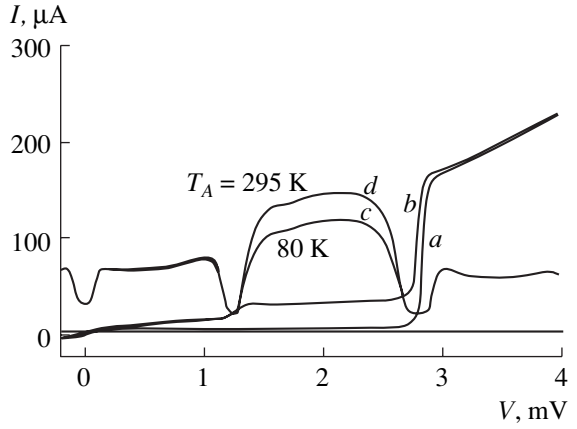


Fig. 4. (*a*) Autonomous and (*b*) pumped *I–V* characteristics of the SIS mixer at 955 GHz and the response to (*c*) "cold" and (*d*) "hot" antenna loads at IF = 1.5 GHz.

high losses near 1 THz due to a lower gap frequency [20].

Two identical mixers were placed 500 μm apart in the central part of a 2 × 2-mm 300-μm-thick high-resistivity silicon substrate, which was mounted on the flat back surface of a silicon elliptic lens at its focus. Such dimensions of the chip cut a 120°-wide portion from the pattern of the planar antenna, beyond which the signal is negligibly small. The second mixer on the same chip is spare. The lens diameter, 10 mm, specified both the diffraction divergence of the beam at 1 THz (about 2°) and the chip mount accuracy (about 10 μm). The IF matching circuit was designed so as to keep the reflection losses below –10 dB over the IF range of 4–8 GHz with the dynamic resistance of the mixer varying within R_d = 50–150 Ω.

DESIGN

The mixer was designed for two frequency ranges: 800–960 and 960–1120 GHz. The integrated structure was optimized individually for each of the bands separately. The double-dipole antenna shown in Fig. 2 is a scaled copy of the antenna of a 500-GHz integrated receiver [16]. The overall dimension of the antenna is 34 × 40 μm, and its elements are 4 μm wide. Initially, we assumed that the low density of the microwave current on the surface of the convergent reflector film cannot cause a significant signal loss and allows the use of almost any metal, for example, Nb. The convergent reflector chip measures 0.5 × 0.5 mm. At about 1 THz, its thickness, 22 μm, equals a quarter of the wavelength in single-crystal silicon (ε = 11.7) coated with a 200-nm-thick Nb film on one side. The antireflection coating of the microwave lens, a 46-μm-thick layer of Stycast™-1264 (ε ≈ 2.9) epoxy compound, is optimized for 960 GHz.

Two Nb/Al/AlO_x/Nb SIS junctions of area 1 μm^2 each are integrated with the antenna as shown in Figs. 1 and 2. The 3-μm-wide transmission line, which connects the antenna and two junctions placed 3.6 or 3.0 μm (depending on the frequency range) apart, is a microstrip line with a narrow (4-μm-wide) ground plane. The mixer was designed under the following assumptions. Since the configuration of the transmission line is almost symmetric, the strip and its screen can be considered interchangeable; therefore, they are connected to the antennas symmetrically. Because of the symmetry of the structure, its center can be regarded as a virtual grounding point in the middle between the two SIS junctions. Thus, the segment of the microstrip line that connects the junctions can be viewed as two independent short-circuited stubs, which are equivalent to two high-frequency tuning inductors parallel-connected to either of the SIS junctions [10].

The match of the antenna and the efficiency of transmitting the signal to the SIS junctions were computed with allowance for ohmic losses in the tuning elements. As shown in Fig. 3, the results computed are in good agreement with the experimental data obtained with a Fourier spectrometer. Our simulations assumed that the NbTiN base electrode is a perfect superconductor with a 300-nm London penetration depth. For the aluminum strip, the sheet resistance was assumed to be 0.15 Ω. Band-stop filters *F*1 and *F*2 are connected to both dipoles as shown in Fig. 1 to make the radiation pattern of the antenna system as symmetric as possible. Note that, as usual, only one filter is connected to the IF channel. The filters are designed as a periodic structure composed of quarter-wavelength sections of coplanar and microstrip lines in order to provide a high performance while retaining the small size of the device. The

ohmic loss in the antenna filters turned out to be appreciable. The attenuation in the three-section filter was estimated at 6%. Basically, this value can slightly be reduced by adding sections; however, this increases the filter capacitance, making the IF matching of the mixer difficult. The curves in Fig. 3 are obtained with regard for unwanted effects, such as the inductance of the SIS junction and the inductance of current spreading around the junction window. The current spreading inductance was estimated at 0.1–0.2 pH [21].

EXPERIMENTAL RESULTS

Pilot mixers were fabricated by the process used to make waveguide devices with NbTiN and Al tuning circuits [20]. We applied conventional optical lithography. The NbTiN ground plane of the microstrip line was 300 nm thick. It was deposited at room temperature. The thickness of the SiO_2 insulator was 250 nm. The conductivity of the 400-nm-thick aluminum layer was $2 \times 10^8\ \Omega^{-1}\ m^{-1}$ at 4 K, which most likely corresponds to the anomalous limit [22]. The aluminum layer is protected against environmental attack by a 200-nm-thick SiO_2 film. A typical *I–V* characteristic of the pilot SIS mixer is shown in Fig. 4.

The parameters of several devices designed for frequency range 4 (960–1120 GHz) were measured in a standard vacuum cryostat with optimized IR filters with losses of 1.2 dB at 1 THz [23]. The Fourier spectrometer data suggested that the frequency response has a feature very much like the high-frequency cutoff near 1 THz. This cutoff frequency is almost independent of the SIS junction dimensions and other important tuning parameters. The double-sideband noise temperature T_{RX} for several mixers was about 500 K above 950 GHz. Since the noise temperature decreased with increasing frequency, we inferred that the device was tuned to a lower-than-optimal frequency. Eventually, a device with a center frequency slightly below 1 THz was selected. The 3-dB frequency band of this mixer determined with the Fourier spectrometer was 800–1050 GHz (Fig. 3). The responses to the change in the temperature of the matched antenna load were 1.6 dB at a mixer temperature of 4.2 K and 1.8 dB at 2 K for 935 GHz in the case of a 15-μm-thick mylar diplexer (Fig. 4). With a thinner diplexer (6 μm), the response increased to 2.1 dB, which corresponds to the double-sideband noise temperature T_{RX} = 260 K. With the correction for reflection, both diplexers gave a noise temperature of about 245 K. A water-vapor absorption line near 990 GHz is clearly seen in Fig. 5; these data were obtained at IF = 1.5 GHz.

Figure 6 shows the reflection coefficient at the IF output calculated as a function of the mixer's dynamic resistance for a total capacitance of the structure (including SIS junctions, microstrip lines, and antenna filters) of ≈0.5 pF. As follows from Fig. 6, the reflection coefficient can be kept below –10 dB for most of the

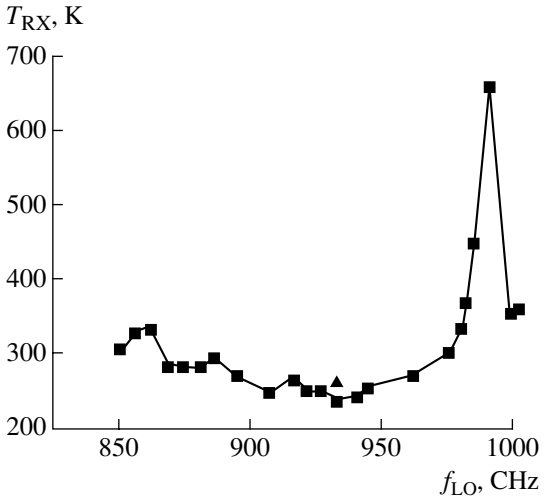


Fig. 5. Noise temperature T_{RX} in the double-sideband mode versus frequency f_{LO} of the local oscillator. The values of T_{RX} measured at about 2 K and corrected for reflection from the 15-μm-thick diplexer (squares). The triangle is a reference data point measured for the 6-μm-thick diplexer (uncorrected).

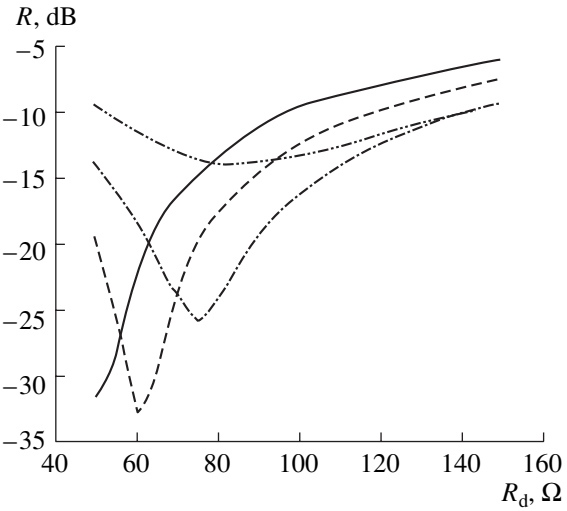


Fig. 6. Reflection losses at the IF output versus dynamic resistance of the mixer with a tuning inductance $L1 = 0.7$ nH (Fig. 1) computed at a frequency of 1.5 (solid line), 4 (dashed line), 6 (dash-and-dot line), and 8 GHz (dash-and-double-dot line).

4–8 GHz range. For a 50-Ω transmission line, this match can be achieved by using a series-connected inductance (Fig. 1). Such an inductance can be provided by bonding wires on the mixer chip. According to our calculations, the required inductance of $L = 0.7$ nH can be obtained with 1-mm-long wires that are 20–50 μm in diameter and placed 0.2 mm apart.

The antenna pattern (Fig. 7) was recorded with a narrow-beam monochromatic source with the mixer operating as a demodulator. Under such conditions, a

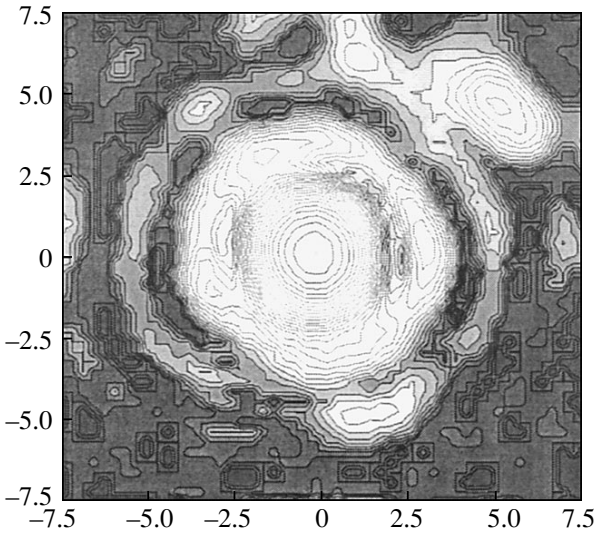


Fig. 7. Antenna pattern at 915 GHz. On both coordinate axes, the angle is plotted in degrees. The intensity isolines are spaced at 1-dB intervals. The darker the color, the lower the intensity. All sidelobes are below –16 dB. The spot in the upper right corner is most likely spurious reflections in the radiator–antenna system.

40-dB dynamic range was achieved. The receiver was turned about the phase center of the antenna. The cross section of the main lobe at 915 GHz is almost circular. The –11-dB half-width of the beam is about 1.8°; the first-order sidelobes are no higher than –16 dB. The quality of the beam remains the same up to 850 GHz. At higher frequencies, the beamwidth is smaller, as could be expected for an optical system operating in the diffraction limit. For the beamwidth to meet the HIFI requirements (*F*/3–*F*/5), we propose to supply the device with an ellipsoidal correcting mirror. The response to the cross-polarized signal component at 850 GHz is no higher than –24 dB relative to the maximum signal. This is an experimental corroboration that a lens combined with a double-dipole antenna is polarized.

CONCLUSION

A terahertz heterodyne receiver built around a quasi-optical Nb/Al/AlO_x/Nb superconducting SIS mixer integrated with a double-dipole antenna is designed, fabricated, and experimentally studied for the first time. The noise temperature of the receiver is equivalent to as little as ten photons. The device covers the 800–1000 GHz range, demonstrating a cutoff frequency of slightly above 1 THz. This result allows us to conclude that a NbTiN/SiO_2/Al tuning microstrip line deposited at room temperature may have relatively low losses at least up to 1 THz. The double-sideband noise temperature was about 250 K at 935 GHz and 360 K at 1 THz, which are the best values achieved to this point in this frequency range. The experimental results are consistent with the computed value of the effective sheet resistance of the NbTiN/SiO_2/Al microstrip line (0.1–0.2 Ω). The simulations show that a twin-type SIS mixer with a double-dipole antenna can be matched at IF in the range of 4–8 GHz. A double-dipole antenna combined with an elliptic silicon lens efficiently receives the signal in the terahertz range. The sidelobe level is below –16 dB; the cross-polar pattern is below –24 dB, which shows that a lens antenna combined with a double-dipole antenna is polarized.

ACKNOWLEDGMENTS

The authors thank Th. De Graauw and V.P. Koshelets for their encouragement and interest in this work and to D. Nguen for technical assistance.

This work was supported by the European Space Agency under the contract ESTEC no. 11653/95, the Russian State Scientific Program "Superconductivity," the Russian Foundation for Basic Research (project no. 00-02-16270), INTAS (grant no. 97-1712), and ISTC (grant no. 1199).

REFERENCES

1. http://www.sron.nl
2. J. R. Tucker and M. J. Feldman, Rev. Mod. Phys. **57**, 1055 (1985).
3. J. W. Kooi, M. Chan, B. Bumble, *et al.*, Int. J. Infrared Millim. Waves **16**, 2049 (1995).
4. A. Karpov, J. Blondel, M. Voss, *et al.*, IEEE Trans. Appl. Supercond. **9**, 4456 (1999).
5. J. A. Stern, B. Bumble, H. G. LeDuc, *et al.*, in *Proceedings of the 9th International Symposium on Space Terahertz Technology, 1998*, p. 305.
6. D. C. Mattis and J. Bardeen, Phys. Rev. **111**, 412 (1958).
7. P. Dieleman, T. M. Klapwijk, J.-R. Gao, *et al.*, IEEE Trans. Appl. Supercond. **7**, 2566 (1997).
8. H. Van de Stadt, A. M. Baryshev, P. Dieleman, *et al.*, in *Proceedings of the 6th International Symposium on Space Terahertz Technology, 1995*, p. 66.
9. V. Yu. Belitsky, S. W. Jacobsson, L. V. Filippenko, *et al.*, in *Proceedings of the 4th International Symposium on Space Terahertz Technology, 1993*, p. 538.
10. J. Zmuidzinas, H. G. LeDuc, J. A. Stern, *et al.*, IEEE Trans. Microwave Theory Tech. **42**, 698 (1994); M. Bin, M. C. Gaidis, J. Zmuidzinas, *et al.*, Appl. Phys. Lett. **68**, 1714 (1996).
11. V. Yu. Belitsky and E. L. Kollberg, J. Appl. Phys. **80**, 4741 (1996); Y. Uzawa, Z. Wang, A. Kawakami, *et al.*, in *Proceedings of the 12th International Symposium on Space Terahertz Technology, San Diego, 2001*.
12. N. N. Iosad, V. V. Roddatis, S. N. Polyakov, *et al.*, IEEE Trans. Appl. Supercond. **11**, 3970 (2001).
13. D. F. Filipovic, S. S. Gearhart, and G. M. Rebeiz, IEEE Trans. Microwave Theory Tech. **14**, 1738 (1993).
14. M. J. M. van der Vorst, P. J. I. De Maagt, and M. H. A. J. Herben, in *Proceedings of the International Symposium on Antennas (JINA'96), 1996*, p. 511;

M. J. M. van der Vorst, PILRAP, Software for Design of Integrated Lens Antennas.

15. A. Skalare, Th. De Graauw, and H. van de Stadt, Microwave Opt. Technol. Lett. **4**, 9 (1991).
16. S. V. Shitov, V. P. Koshelets, A. B. Ermakov, *et al.*, IEEE Trans. Appl. Supercond. **9**, 3773 (1999).
17. J. Zmuidzinas and H. G. LeDuc, IEEE Trans. Microwave Theory Tech. **40**, 1797 (1992).
18. S. V. Shitov, A. M. Baryshev, V. P. Koshelets, *et al.*, in *Proceedings of the 7th International Symposium on Space Terahertz Technology, 1996*, p. 525.
19. B. D. Jackson, N. N. Iosad, B. Leone, *et al.*, in *Proceedings of the 10th International Symposium on Space Terahertz Technology, 1999*, p. 144; B. Leone, B. D. Jackson, J.-R. Gao, *et al.*, Appl. Phys. Lett. **76**, 780 (2000).
20. B. D. Jackson, G. De Lange, W. M. Laauwen, *et al.*, in *Proceedings of the 11th International Symposium on Space Terahertz Technology, 2000*, p. 238.
21. M. M. Khapaev, Supercond. Sci. Technol. **9**, 729 (1996); M. Yu. Kupriyanov, private communication of the inductance calculation.
22. R. L. Kautz, J. Res. Natl. Bur. Stand. **84**, 247 (1979).
23. A. M. Baryshev, B. D. Jackson, G. De Lange, *et al.*, in *Proceedings of the 11th International Symposium on Space Terahertz Technology, 2000*, p. 129; B. D. Jackson, A. M. Baryshev, G. De Lange, *et al.*, in *Proceedings of the 12th International Symposium on Space Terahertz Technology, San Diego, 2001*.

Translated by A. Khzmalyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1158–1165. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 93–101.

ACOUSTICS, ACOUSTOELECTRONICS

Synthesis and Analysis of SAW Dispersive Ladder Transducers by the Modified Method of Coupled Modes

V. F. Dmitriev

Avangard–Élionika Joint-Stock Company, St. Petersburg, 195271 Russia

e-mail: elionica@cityline.spb.ru

Received January 31, 2002

Abstract—A method for the synthesis and analysis of SAW filters and delay lines that use dispersive ladder interdigital transducers is described. A method for designing SAW devices that relies on modified equations for coupled waves is proposed. A singular integral equation for the surface current density in the transducer's electrodes is derived and solved. Theoretical data are compared with experimental results. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Dispersive interdigital transducers (IDTs) in which the center line of the electrode apertures is inclined to the direction of SAW propagation (ladder-type IDTs) [1] are widely employed in SAW devices. The main purpose of using ladder geometry in dispersive IDTs is to suppress second-order effects in the device's frequency response. In ladder geometry, the regions of excitation and propagation of SAWs with different frequencies can be spatially separated. The electrodes that do not excite an SAW of a particular frequency are displaced from the SAW propagation path; accordingly, the number of electrodes in each cross section of an IDT decreases (Fig. 1). As a result, second-order effects in the amplitude–frequency response (AFR) and the phase–frequency response (PFR) of the device are suppressed. The inclination of the electrode center line depends on the electrode arrangement in an IDT and on a desired electrical performance of the device.

Dispersive ladder IDTs can be used in dispersive delay lines with a quadratic PFR and in filters with a linear PFR. The shape of the characteristic (linear or quadratic) depends on the slopes of the dispersive characteristics of the input and output IDTs. When the signs of these slopes are the same, the device operates as a filter (Fig. 1) or a nondispersive delay line and has a linear PFR. When the slopes are of opposite sign, the device operates as a dispersive delay line (Fig. 1) and exhibits a quadratic PFR.

The synthesis method for ladder IDTs was proposed in [1], and the method of analysis based on a physical model and on partitioning the layout into channels was reported in [2]. The key idea in the synthesis of a ladder IDT is that the optimal number of electrodes in each IDT cross section must equal the effective number of electrodes: $N_{\text{eff}}(f) = f[T/\Delta f]^{1/2}$, where f is the synchronism frequency, T is the IDT delay due to dispersion, and Δf is the operating bandwidth. However, approximate relationships for the synthesis of SAW devices [1] are often far from accurate. Also, the physical model sometimes does not provide a reliable procedure for calculating the AFR from a given IDT layout. Therefore, at the design stage, one may fail in properly choosing an IDT layout that provides a flat (or another given) AFR, and it becomes necessary to correct the layout after fabricating the device. The physical model also neglects the effect of three-transit signals and reflections from the electrodes on the AFR and PFR.

In this paper, we propose modified equations for coupled modes (COM equations) that describe a wide class of SAW devices: filters, resonators, and dispersive and nondispersive delay lines. The modified COM equations are central to a method for the synthesis and analysis of SAW filters and dispersive delay lines that employ dispersive ladder IDTs. A singular integral equation for the surface current density on the IDT

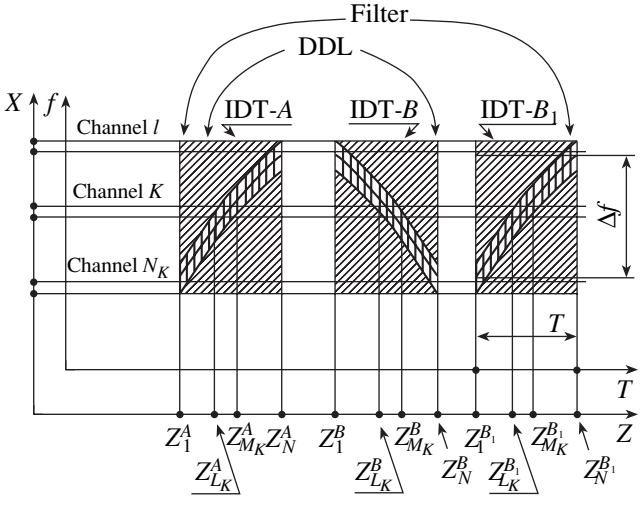


Fig. 1. Geometry of a dispersive delay line (DDL) and filter.

1063-7842/02/4709-1158$22.00 © 2002 MAIK "Nauka/Interperiodica"

electrodes that includes edge effects, the finiteness of the IDT length, the effect of the acoustic field, and the variation of the acoustic wave amplitude under the electrodes is derived and solved. The designs of a filter and a dispersive delay line based on dispersive ladder IDTs are presented. The theoretical results are compared with experimental data.

SYNTHESIS OF SAW DEVICES BUILT ON DISPERSIVE LADDER IDTs

Let it be necessary to design a SAW device with a ladder IDT where losses decrease or increase linearly with frequency. We will use a method similar to that reported in [1]. In this case, the power emitted by the IDT in a frequency range of interest is given by

$$U_0^2 G_A(f) = P_0 K_0(f), \tag{1}$$

where $G_A(f)$ is the IDT admittance, P_0 is the power emitted by the IDT at the center frequency f_0, $K_0(f) = 1 + K_s[f - f_0]/\Delta f$, $U_0 = \text{const}(f)$ is the amplitude of the input signal, and $K_s = \text{const}(f)$ is the slope of the frequency response.

For the insertion losses to decrease linearly with increasing frequency, it is necessary that $K_s > 0$; for the losses to increase linearly with frequency, $K_s < 0$. Manipulations similar to those used in [1] yield the following frequency dependence of the IDT aperture:

$$W(f) = W(f_0)[f_0/f]K_0(f) \times \{[(R_G)^2 + X^2(f)]/[(R_G)^2 + X^2(f_0)]\}^{1/2}, \tag{2}$$

where $W(f_0)$ is the IDT aperture at the center frequency, which is chosen so as to suppress diffraction effects and provide reasonable insertion losses, R_{sg} is the impedance of the signal generator, and $X(f)$ is the imaginary (reactive) part of the IDT input impedance.

The position of the center line of the electrode apertures (or the coordinates of their centers) will be found from the condition that the IDT dimension in the transverse direction is independent of the X coordinate: $L_{\text{eff}}(X) = N_{\text{eff}}(f)\lambda/2 = \text{const}$ [1]. Note that each of the X coordinates is related to a particular frequency f within the passband. From Fig. 2,

$$\tan\phi_1 = W_i/L_{\text{eff}}, \tag{3}$$

$$\tan\phi_2 = \Delta X_i/\Delta Z_i, \tag{4}$$

where $\Delta X_i = X_{i+1} - X_i$ and $\Delta Z_i = Z_{i+1} - Z_i$ are the center distances of the ith and $(i + 1)$th electrodes in the X and Z coordinates, respectively; W_i is the aperture of the ith electrode; and ϕ_1 and ϕ_2 are, respectively, the angles made by the aperture boundary and the electrode center line with the Z axis (Fig. 2).

If the IDT aperture varies slowly ($(W_{i+1} - W_i)/W_{i+1} \ll 1$), so does the curvature of the center line.

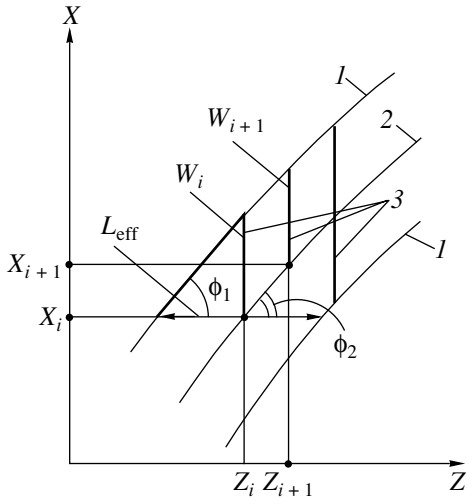


Fig. 2. Detail of the ladder transducer: (*1*) boundaries of the transducer electrode apertures, (*2*) line passing through electrode centers, and (*3*) electrodes.

In this case, $\phi_1 \approx \phi_2$. Then, formulas (3) and (4) yield

$$\Delta X_i \approx \Delta Z_i W_i/L_{\text{eff}} \tag{5}$$

with $X_{i+1} = X_i + \Delta X_i$ and $X_1 = 0$.

Next, we find the Z coordinates of the electrode centers taking into account metallizations in the ladder IDT, as for IDTs of linear layout:

$$Z_i = B_1\{[1 + B_2(I - 1)]^{1/2} - 1\}, \tag{6}$$

where $B_1 = \pm V f_N T_\infty/2$, $B_2 = \pm\Delta f_\infty/[(f_N)^2 T_\infty]$, and T_∞ and Δf_∞ are the total delay due to dispersion and the bandwidth of the IDT, respectively. The "–" sign refers to the positive slope of the dispersion characteristic with $f_N = f_0 + \Delta f_\infty/2$, the "+" sign, to the negative slope of the dispersion characteristic with $f_N = f_0 - \Delta f_\infty/2$; i runs from 1 to N.

Note that $T_\infty > T$, $\Delta f_\infty > \Delta f$, and $\Delta f_\infty/T_\infty = \Delta f/T$.

MODIFIED COM METHOD

The usual COM theory (see, e.g., [3]), based on a set of inhomogeneous differential equations, greatly complicates the design of SAW devices. In the framework of this theory, it is difficult to consider such factors as the variable period of the structure, apodization, and the nonuniform distribution of the surface charge over the electrodes. It is sufficiently easy to include all these factors by using the modified COM method, which deals with an elementary component of the structure (an electrode of the interdigital transducer or of the reflecting structure). This method can also be readily extended to a more complex model of the structure. Here, the parameters of the SAW structure as a whole (interdigital transducer, reflecting structure, or their arbitrary combination) are found by multiplying the correspond-

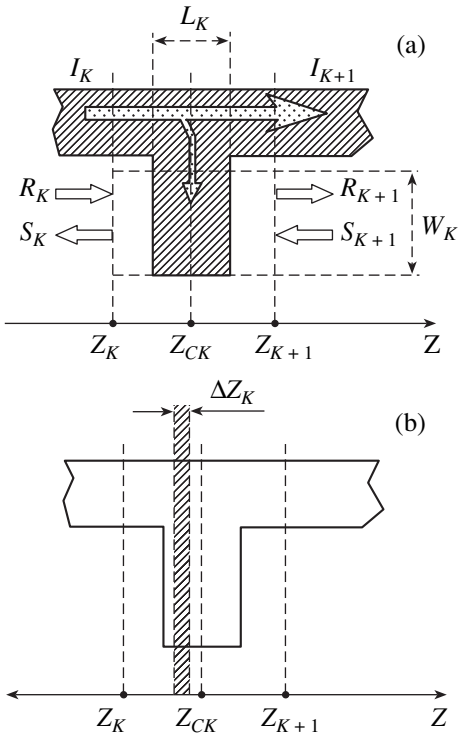


Fig. 3. *K*th electrode of the transducer.

ing *P* matrices for individual electrodes (as is done in the conventional theory of quadripoles in terms of scattering matrices).

Let an SAW structure consist of variously spaced alternating-polarity electrodes. We assume that adjacent electrodes may overlap and that a source with the signal amplitude U_0 is placed on the left. Consider the *K*th electrode of the IDT (Fig. 3a). Let $R(Z, \omega)$ and $S(Z, \omega)$ be coupled uniform plane waves propagating through the electrode structure of the transducer in the positive and negative *Z* directions, respectively. The expressions for our uniform plane waves are

$$R(Z, \omega) = R(\omega)\exp(-j\kappa Z), \quad (7)$$

$$S(Z, \omega) = S(\omega)\exp(+j\kappa Z), \quad (8)$$

where $R(\omega)$ and $S(\omega)$ are the complex amplitudes of the corresponding waves and κ is the wave number.

Let a wave $R_K(Z, \omega)$ be incident on the *K*th electrode from the left; the wave $S_{K+1}(Z, \omega)$, from the right. Then, with allowance for reflection, transmission, and transformation with a coefficient ξ_K, the complex amplitudes of the transmitted waves can be written as

$$\begin{aligned} S_K(\omega) &= r_K\eta_{1K}\exp[-j(\kappa_{\text{eff}} - \kappa_0)p_K]R_K(\omega) \\ &+ \eta_{1K}(1 - |r_K|^2)^{1/2}\exp[-j(\kappa_{\text{eff}} - \kappa_0)p_K]S_{K+1}(\omega) \\ &+ \xi_K(\omega)\eta_{2K}\exp[-j(\kappa_{\text{eff}} - \kappa_0)p_K/2]U_0, \end{aligned} \quad (9)$$

$$\begin{aligned} R_{K+1}(\omega) &= \eta_{1K}(1 - |r_K|^2)^{1/2}\exp[-j(\kappa_{\text{eff}} - \kappa_0)p_K] \\ &\times R_K(\omega) + r_K\eta_{1K}\exp[-j(\kappa_{\text{eff}} - \kappa_0)p_K]S_{K+1}(\omega) \\ &+ \xi_K(\omega)\eta_{2K}\exp[-j(\kappa_{\text{eff}} - \kappa_0)p_K/2]U_0, \end{aligned} \quad (10)$$

where r_K is the complex coefficient of reflection from the *K*th electrode, κ_{eff} is the effective wave number of the SAW, $\kappa_0 = 2\pi/p_K$, $p_K = Z_{K+1} - Z_K$, ξ_K is the SAW transformation coefficient at the *K*th electrode, $\eta_{1K} = W_{1K}/W_0$, $\eta_{2K} = W_{2K}/W_0$, W_0 is the maximal aperture, and W_{1K} is the overlap of electrodes. $W_{2K} = W_0$ when idle electrodes are used; otherwise, $W_{2K} = W_{1K}$.

The phase factors at the terms standing for wave reflection (transformation) define the phase shift the SAW acquires when traveling from the center of reflection (transformation) to the corresponding boundary (Z_K for $S_K(\omega)$ and Z_{K+1} for $R_K(\omega)$). The center of reflection (transformation) is assumed to be at the center of the electrode. The effective wave number is found as $\kappa_{\text{eff}} = 2\pi/\lambda_E = \omega/[V_0 + L_K(V_m - V_0)/p_K] - j\alpha$, where V_0 is the SAW velocity on the free surface, V_m is the SAW velocity under the metallized surface, and α are the total losses of the SAW per unit length.

The current in the IDT varies because of the transformation of forward and backward waves and a voltage drop across the electrode capacitance:

$$\begin{aligned} \Delta I_K(\omega) &= I_K(\omega) - I_{K+1}(\omega) = +2\xi_K(\omega) \\ &\times \exp[-j(\kappa_{\text{eff}} - \kappa_0)p_K/2]R_K(\omega) + 2\xi_K(\omega) \\ &\times \exp[-j(\kappa_{\text{eff}} - \kappa_0)p_K/2]S_K(\omega) + j\omega(C_2/2)U_0. \end{aligned} \quad (11)$$

Consider the terms standing for SAW transformation as the wave passes through the transducer electrode (Fig. 3b) in view of the fact that the excitation is distributed. Let the efficiencies of direct and inverse SAW transformations on the electrodes be equal; i.e., the transformation obeys the reciprocity principle. The surface current distribution on the electrodes, $J(Z)$, is assumed to be given. Assume also that a small region ΔZ_K of the electrode transforms the SAW in the same manner as the electrode as a whole. Then, we add up the contributions to the SAW transformation over the electrode width relative to its center Z_c and pass to the limit ($\Delta Z_K \longrightarrow 0$) to obtain

$$\xi_K = G_a \int_{-L_K/2}^{L_K/2} J(Z)\exp[-j(\omega/V_m - \kappa_0)Z]dZ, \quad (12)$$

where G_a is the radiation acoustic admittance at the synchronism frequency (see, e.g., [3]).

The distribution $J(Z)$ of the surface current over the electrodes is calculated by a self-consistent technique, i.e., with allowance for edge effects, the finiteness of the length of the transducer, and the backward reaction of the piezoelectric material (see the next section).

Formulas (9)–(11) can be written in matrix form:

$$\begin{vmatrix} S_K(\omega) \\ R_{K+1}(\omega) \\ \Delta I_K(\omega) \end{vmatrix} = \begin{vmatrix} P(1,1) & P(1,2) & P(1,3) \\ P(2,1) & P(2,2) & P(2,3) \\ P(3,1) & P(3,2) & P(3,3) \end{vmatrix} \begin{vmatrix} R_K(\omega) \\ S_{K+1}(\omega) \\ U_0 \end{vmatrix}. \quad (13)$$

The P matrix of an interdigital transducer as a whole can be calculated by multiplying the P matrices for each of the electrodes. Applying Eqs. (9)–(11) with arbitrary coefficients to two SAW structures connected in series, we easily find the components of the complete P matrix:

$$P^{(c)}(1,1) = P^{(1)}(1,1) + P^{(1)}(1,2)P^{(2)}(1,1)P^{(1)}(2,1)/P^{(0)}, \quad (14)$$

$$P^{(c)}(1,2) = P^{(1)}(1,2)P^{(2)}(1,2)/P^{(0)}, \quad (15)$$

$$P^{(c)}(1,3) = P^{(1)}(1,3) + P^{(1)}(1,2)[P^{(2)}(1,3) + P^{(2)}(1,1)P^{(1)}(2,3)]/P^{(0)}, \quad (16)$$

$$P^{(c)}(2,1) = P^{(1)}(2,1)P^{(2)}(2,1)/P^{(0)}, \quad (17)$$

$$P^{(c)}(2,2) = P^{(2)}(2,2) + P^{(2)}(2,1)P^{(1)}(2,2)P^{(2)}(1,2)/P^{(0)}, \quad (18)$$

$$P^{(c)}(2,3) = P^{(2)}(2,3) + P^{(2)}(2,1)[P^{(1)}(2,3) + P^{(2)}(1,3)P^{(1)}(2,2)]/P^{(0)}, \quad (19)$$

$$P^{(c)}(3,1) = P^{(1)}(3,1) + P^{(1)}(2,1)[P^{(1)}(3,1) + P^{(2)}(1,1)P^{(1)}(3,2)]/P^{(0)}, \quad (20)$$

$$P^{(c)}(3,2) = P^{(2)}(3,2) + P^{(2)}(1,2)[P^{(1)}(3,2) + P^{(1)}(2,2)P^{(2)}(3,1)]/P^{(0)}, \quad (21)$$

$$P^{(c)}(3,3) = P^{(1)}(3,3) + P^{(2)}(3,3) + \{P^{(1)}(3,2)[P^{(2)}(1,3) + P^{(2)}(1,1)P^{(1)}(2,3)] + P^{(2)}(3,1)[P^{(1)}(2,3) + P^{(1)}(2,2)P^{(2)}(3,1)]\}/P^{(0)}, \quad (22)$$

where $P^{(0)} = 1 - P^{(2)}(1,1)P^{(1)}(2,2)$. The superscripts "c," 1, and 2 refer to the complete P matrix, P matrix of the SAW structure on the left, and P matrix of the SAW structure on the right, respectively. By an SAW structure, we mean both an individual electrode and a group of electrodes for which the P matrix has been calculated.

The expressions presented above can be used for calculating the input admittance of an interdigital transducer incorporated into a filter or a resonator with arbitrarily varying electrode spacings and electrode apertures and with the actual distribution of the surface current (charge) over the electrodes. Note that the input admittance of the transducer is specified by the element $P^{(c)}(3,3)$ of the P matrix for the whole SAW structure.

SURFACE CURRENT DISTRIBUTION OVER THE ELECTRODES

Let the complex amplitudes $R_f(\omega)$ and $S_f(\omega)$ (f is the electrode number) of surface acoustic waves under each of the electrodes be known. We will derive an equation for $J(Z)$ using the electrodynamic boundary condition for the tangential component of an ac electric field on the surface of a perfectly conducting infinitely thin metal sheet (IDT):

$$E^I(Y = 0, Z) + E^A(Y = 0, Z) = 0, \quad (23)$$

where $E^I(Y = 0, Z)$ is the tangential component of the electric field produced by a system of strip currents on the electrode surfaces, and $E^A(Y = 0, Z)$ is the tangential component of the electric field due to the surface acoustic waves $R(Z, \omega)$ and $S(Z, \omega)$ that propagate under the electrodes on the piezoelectric surface.

Boundary condition (23) yields

$$\eta \int_{\mathcal{L}} dZ' J(Z')/(Z' - Z) - R_f(\kappa_{\text{eff}})\exp[-j(\kappa_{\text{eff}} - \kappa_0)Z] + S_f(\kappa_{\text{eff}})\exp[j(\kappa_{\text{eff}} - \kappa_0)Z] = 0, \quad (24)$$

where the integration path $\mathcal{L}$ consists of intervals $[a_f, b_f]$, a_f and b_f are the coordinates of the electrode edges on the Z axis, $\eta = [\pi\varepsilon_\infty(\kappa_{\text{eff}} - \kappa_0)]^{-1}$, ε_∞ is the permittivity of the piezoelectric, and $R_f(\kappa_{\text{eff}})$ and $S_f(\kappa_{\text{eff}})$ are the complex amplitudes of the forward and backward surface acoustic waves traveling in the $\pm Z$ directions under electrode no. f.

Assume that the complex amplitudes of the potentials remain constant within a particular electrode but may vary from electrode to electrode in accordance with Eqs. (9)–(11).

Equation (24) is a singular integral equation that describes the surface current distribution over the IDT electrodes with allowance for the reaction of the excited wave. To solve this equation, we will take advantage of the method applied earlier [4, 5] to the integral equation for current distribution in spin-wave transducers. First, we apply the inversion procedure to the Cauchy integral [6] to remove the singularity:

$$J(z) = \left[\prod_{f = L_K^A}^{M_K^A} (z - a'_f)(b'_f - z)\right]^{-1/2} \{Q_{m-1}(z) + \int_{\mathcal{L}} dz' J(z')[F^+(z', k)\exp(-jkz') + F^-(z', k)\exp(jkz')]\} = 0, \quad (25)$$

where $a'_f = 2a_f/(b_1 - a_1)$, $b'_f = 2b_f/(b_1 - a_1)$, $z = 2Z/(b_1 - a_1)$, $k = (\kappa_{\text{eff}} - \kappa_0)(b_1 - a_1)$, and $Q_{m-1}(Z) = C_{m-1}z^{m-1} + C_{m-2}z^{m-2} + \ldots C_0$ is an algebraic polynomial of degree no higher than $m - 1$ whose coefficients are found from the condition imposed on the magnitude and direction of the currents in the electrodes:

$$I_f = \int_{a_f}^{b_f} dZJ(Z).$$

The quantities $F^{\pm}(z, k)$ are defined as

$$F^{\pm}(z', k) = \int_{\mathcal{L}} dz_1 \left\{ \left[\prod_{f = L_K^A}^{M_K^A} (z_1 - a'_f) \right. \right. \times (b'_f - z_1) \left. \left. \right]^{1/2} / (z_1 - z') \right\} \Lambda_f^{\pm}(k) \exp[\pm jkz_1], \tag{26}$$

where $\Lambda_f^+(k) = (\pi\eta)^{-1}R_f(k)$ and $\Lambda_f^-(k) = (\pi\eta)^{-1}S_f(k)$.

On each interval of integration $a_f \le Z \le b_f$, we change the variables to $z'_f = 2z - (b_f + a_f)/(b_1 - a_1)$, coming to $J(z) = [1 - (z_f)^2]^{-1/2}J_1(z_f)$, and then $z'_f = \sin(\tau_f)$. In this way, we demonstrate that the kernels $F^{\pm}(z, k)$ in (26) are Fredholm kernels. Then, (25) is an inhomogeneous Fredholm integral equation of the second kind and its solution has the form [7]

$$J(z) = \left[\prod_{f = L_K^A}^{M_K^A} (z - a'_f)(b'_f - z) \right]^{-1/2} \times [Q_{m-1}(z) + C^+F^+(z, -k) + C^-F^-(z, -k)], \tag{27}$$

where the coefficients $C^{\pm}$ are calculated from the system of equations according to the Cramer rule:

$$\begin{aligned} C^+ - C^+F_1^-(+k) + C^-F_1^+(+k) &= f_1(+k), \\ C^- - C^+F_1^-(-k) + C^-F_1^+(-k) &= f_1(-k), \end{aligned} \tag{28}$$

where

$$f_1(\pm k) = \int_{\mathcal{L}} dz \left\{ \left[\prod_{f = L_K^A}^{M_K^A} (z - a'_f)(b'_f - z) \right]^{-1/2} \times Q_{m-1}(z) \exp[\pm jkz] \right\}, \tag{29}$$

$$F_1^{\pm}(\pm k) = \int_{\mathcal{L}} dz \left\{ \left[\prod_{f = L_K^A}^{M_K^A} (z - a'_f)(b'_f - z) \right]^{-1/2} \times F^{\pm}(k, z) \exp[\pm jkz] \right\}. \tag{30}$$

Expressions (27)–(30) describe the surface current distribution over the electrodes with allowance for their mutual coupling and reaction of the surface acoustic wave.

If the variation of the complex amplitudes $R_f(\kappa_{\text{eff}})$ and $S_f(\kappa_{\text{eff}})$ of the acoustic waves with propagation length can be neglected (as for materials with weak piezoelectric properties), $J(Z)$ can be determined directly from (27) in view of Eqs. (28)–(30). For materials with pronounced piezoelectric properties, the complex amplitudes $R_f(k)$ and $S_f(k)$ vary significantly as the waves propagate through an IDT, and $J(Z)$ should be sought by the method of successive iterations. At the initial stage of analysis, the response of the piezoelectric to the surface current distribution should be neglected. In this case, the zeroth-order approximation $J^0(Z)$ to $J(Z)$ is given by (27) with $C^{\pm} = 0$. Further, having calculated $R_f(k)$ and $S_f(k)$ from (9)–(11) and (14)–(22) for $J(Z) = J^0(Z)$, we find $C^{\pm}$ and obtain the first-order approximation to $J(Z)$ from (27) with allowance for the response of the piezoelectric to the current distribution in the IDT electrodes. By repeating this procedure, any higher order approximation to $J(Z)$ can be calculated.

ANALYSIS OF SAW DEVICES WITH DISPERSIVE IDTs BY THE COM METHOD

Consider an SAW device with a dispersive IDT where the center line of the electrodes has a gradually varying slope relative to the propagation direction (Fig. 1). To calculate the admittance matrix $\hat{Y}$ of the input (output) IDT, we will use the modified COM method (the related equations were derived above) and partition the original SAW structure into channels. The components of the input admittance will then be found by summing over the channels:

$$Y(l_Y, m_Y) = \sum_{K=1}^{N_K} Y_K(l_Y, m_Y), \tag{31}$$

where N_K is the number of channels, $Y_K(l_Y, m_Y)$ are the components of the input admittance matrix for the Kth channel, $l_Y = 1$ or 2, and $m_Y = 1$ or 2.

We assume that the number of channels is sufficiently large; therefore, the analytical frequency

responses do not change with increasing N_K. When calculating the contribution $Y_K(l_Y, m_Y)$ of the Kth channel to the total admittance of the SAW device, we will consider the kth channel as an independent SAW device (Fig. 4) and use the components of the P matrices for the input, $P_K^{(A)}(l_P, m_P)$, and output, $P_K^{(B)}(l_P, m_P)$, IDTs in the Kth channel. Then, for the input admittance of the Kth channel, we have

$$Y_K(1,1) = P_K^{(A)}(3,3) + P_K^{(B)}(1,1)P_K^{(A)}(3,2)P_K^{(A)}(2,3)/Y_{K0}, \quad (32)$$

$$Y_K(1,2) = P_K^{(A)}(3,2)P_K^{(B)}(1,3)\Phi_K/Y_{K0}, \quad (33)$$

$$Y_K(2,1) = P_K^{(A)}(2,3)P_K^{(B)}(3,1)\Phi_K/Y_{K0}, \quad (34)$$

$$Y_K(2,2) = P_K^{(B)}(3,3) + P_K^{(A)}(2,2)P_K^{(B)}(1,3)P_K^{(B)}(3,1)/Y_{K0}, \quad (35)$$

where $Y_{K0} = (\Phi_K)^2 - P_K^{(A)}(2,2)P_K^{(B)}(1,1)$, $\Phi_K = \exp(j2\pi Z_K^{(AB)}/\lambda - \alpha Z_K^{(AB)})$, $Z_K^{(AB)} = Z_1^{(B)} - Z_N^{(A)}$ is the distance between the first electrode of IDT-B and the last electrode of IDT-A (Fig. 1), and λ is the wavelength on the free surface.

The components $P_K^{(A)}(l_P, m_P)$ and $P_K^{(B)}(l_P, m_P)$ will be calculated with the modified COM method. To synchronize the channels, it is necessary to take into account the initial phase of the first electrode for each of the channels in IDT-A (IDT-B). To this end, we introduce phase factors into P-matrix components (13) for the first electrodes of the channels:

$$P^{(A)}(1,1) = P(1,1)[F_1^{(A)}\{Z_{K,1}^{(A)}\}]^2, \quad (36)$$

$$P^{(A)}(1,2) = P(1,2)F_1^{(A)}\{Z_{K,1}^{(A)}\}, \quad (37)$$

$$P^{(A)}(1,3) = P(1,3)F_1^{(A)}\{Z_{K,1}^{(A)}\}, \quad (38)$$

$$P^{(A)}(2,1) = P^{(A)}(1,2), \quad (39)$$

$$P^{(A)}(3,1) = P(3,1)F_1^{(A)}\{Z_{K,1}^{(A)}\}, \quad (40)$$

where $F_1^{(A)}\{Z_{K,1}^{(A)}\} = \exp(j2\pi Z_{K,1}^{(A)}/\lambda_M - \alpha Z_{K,1}^{(A)})$, $Z_{K,1}^{(A)} = Z_{L_K}^{(A)} - Z_1^{(A)}$ is the distance between the first electrode (with the number L_K) in the Kth channel of IDT-A and the first electrode of the whole IDT-A, and λ_m is the wavelength under the metallized surface (contact line). The remaining P-matrix components for the first electrode remain unchanged.

With regard for the phase factors, the P-matrix components for the last electrodes in each of the IDT-A channels take the form

$$P^{(A)}(1,2) = P(1,2)F_2^{(A)}\{Z_{K,2}^{(A)}\}, \quad (41)$$

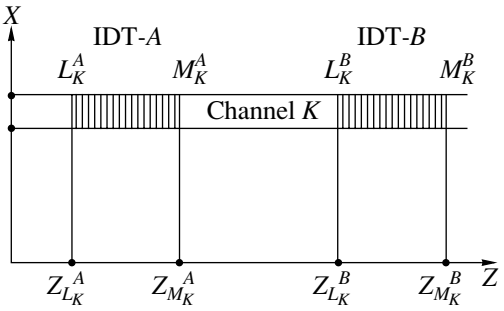


Fig. 4. Kth channel of the device.

$$P^{(A)}(2,1) = P^{(A)}(1,2), \quad (42)$$

$$P^{(A)}(2,2) = P(2,2)[F_2^{(A)}\{Z_{K,2}^{(A)}\}]^2, \quad (43)$$

$$P^{(A)}(2,3) = P(2,3)F_2^{(A)}\{Z_{K,2}^{(A)}\}, \quad (44)$$

$$P^{(A)}(3,2) = P(3,2)F_2^{(A)}\{Z_{K,2}^{(A)}\}, \quad (45)$$

where $F_2^{(A)}(Z_{K,2}^{(A)}) = \exp(j2\pi Z_{K,2}^{(A)}/\lambda_m - \alpha Z_{K,2}^{(A)})$ and $Z_{K,2}^{(A)} = Z_N^{(A)} - Z_{M_K}^{(A)}$ is the distance between the last electrode (with the number N) of the whole IDT and the last electrode of the Kth channel (with the number M_K). The remaining P-matrix components for the last electrode remain unchanged.

Similar expressions can also be written for IDT-B (components $P_K^{(B)}(l, m)$) by setting $F_2^{(B)}(Z_{K,2}^{(B)}) = 1$, because the phase shift of the wave outside IDT-B is insignificant.

Now, the P-matrix components $P_K^{(A)}(l_P, m_P)$ and $P_K^{(B)}(l_P, m_P)$ for the Kth channel can be calculated by multiplying the respective components for the electrodes in the channel:

$$P_K^{(A)}(l_P, m_P) = \prod_{n=L_K^A}^{M_K^A} P_{Kn}^{(A)}(l, m),$$
$$P_K^{(B)}(l_P, m_P) = \prod_{n=L_K^B}^{M_K^B} P_{Kn}^{(B)}(l, m), \quad (46)$$

where L_K and M_K are the numbers of the first and last electrodes in each channel (Fig. 4) and the product sign means the successive calculation of the products according to formulas (14)–(22).

THE DESIGN OF A FILTER WITH DISPERSIVE LADDER IDTs

With the synthesis method proposed above and the analysis based on the modified COM method, we

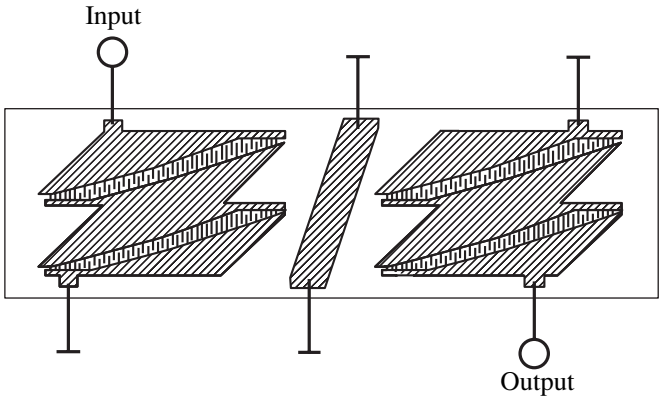


Fig. 5. Geometry of the filter with f_0 = 105 MHz and Δf = 10 MHz.

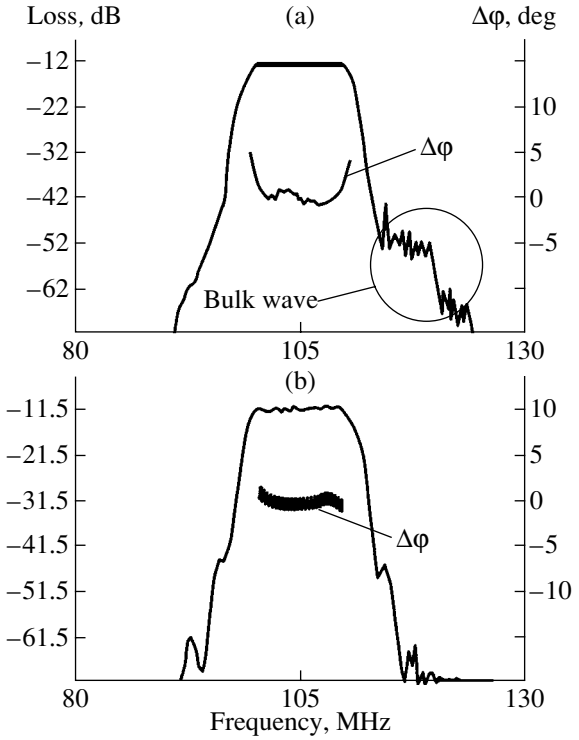


Fig. 6. (a) Experimental frequency responses of the filter and (b) those calculated by the modified COM method.

designed and fabricated a dispersive-IDT-based filter with f_0 = 105 MHz, Δf = 10 MHz, and K_P = 1.45. To increase the input and output impedances, the input and output transducers were composed of two identical IDTs each (Fig. 5). The entire filter circuit occupied a 18 × 4.8-mm board. Based on the theoretical analysis, the dispersion delay per IDT (T_∞) necessary to provide the required bandwidth ratio of the AFR was chosen to be 1.9 μs at Δf_∞ = 15 MHz. To match the input and output of the filter with 50-Ω transmission lines, the electrode aperture at the center frequency, $W(f_0)$, was chosen to be $11\lambda_0$ and a 100-nH matching inductance was employed. The small value of the aperture did not cause diffraction-related distortions in the filter's AFR and PFR, because the piezoelectric material was *Y, Z*-cut $LiNbO_3$. To provide the dispersion delay desired, each of the IDTs consisted of 403 split electrodes. To eliminate Fresnel ripple on the filter's frequency responses, apodization was applied to the first 75 and last 75 electrodes of the IDT. Our calculations showed that K_s equal to 0.3 provides a flat AFR.

The board was fabricated by photolithography and electron sputtering of Al in a vacuum. Experimental frequency responses of the filter are shown in Fig. 6a; those calculated by the modified COM method at N_K = 400, in Fig. 6b. The AFRs were taken with an R4-37 meter of complex transmission coefficients; the PFRs, with a G4-176 precision signal generator and an FK2-12 phase meter. The phase response $\Phi(f)$ is represented in terms of the deviation from the linear law: $\Delta\Phi = \Phi(f) - 2\pi T_0(f_0 - f)$, where T_0 = 2.3179 μs is the filter delay at the frequency f_0. The out-of-band attenuation in the high-frequency range is slightly smaller than the theoretical value because of the spurious excitation of bulk waves. To suppress them, the back surface of the substrate was made corrugated. Without the corrugation, the out-of-band attenuation in the high-frequency range was 10 dB lower than in Fig. 6a. Apparently, refined corrugation would enhance the out-of-band attenuation.

THE DESIGNS OF DISPERSIVE DELAY LINES WITH LADDER-TYPE IDTs

With the synthesis method proposed above and the analysis based on the modified COM method, we designed pulse-forming (with an unweighted IDT aperture) and pulse-compressing (with a weighted IDT aperture) dispersive delay lines (DDLs) with f_0 = 700 MHz, Δf = 200 MHz, and T = 0.6 μs. The analysis showed that, in the frequency range of interest, the flat-top AFR of the pulse-forming DDL is achieved at K_s = 0, Δf_∞ = 260 MHz, and T_∞ = 0.34 μs for each of the IDTs. The electrode aperture at the center frequency, $W(f_0)$, was chosen to be $25\lambda_0$. As the piezoelectric material, we used *Y, Z*-cut $LiNbO_3$. The dispersion delay required was provided with 549 solid electrodes in each of the IDTs. To eliminate Fresnel ripples in the frequency responses of the pulse-forming DDL, the first 64 and the last 46 IDT electrodes were apodized. The Taylor function with a –45-dB sidelobe level was used as a weighting function for the compressing DDL. The delay line boards measured 13 × 3.5 mm.

The delay lines were fabricated by lift-off lithography, electron-beam sputtering of Al in a vacuum, and ion–chemical etching of the piezoelectric surface. The minimal wavelength of the wave excited by the IDT was ≈4 μm, and the electrodes were ≈1 μm wide. Therefore, the resolution of optical lithography was insufficient for split electrodes to be used. To mitigate

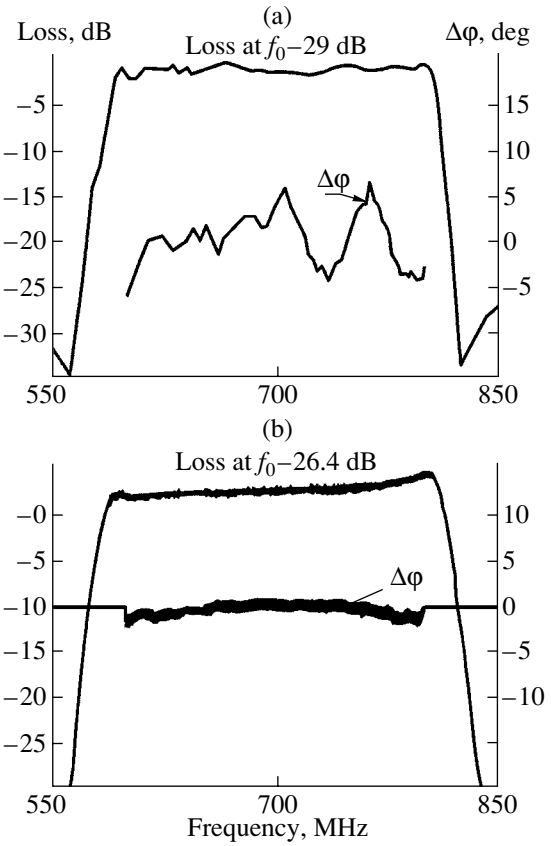


Fig. 7. (a) Experimental frequency responses of the pulse-expansion dispersive delay line and (b) those calculated by the modified COM method.

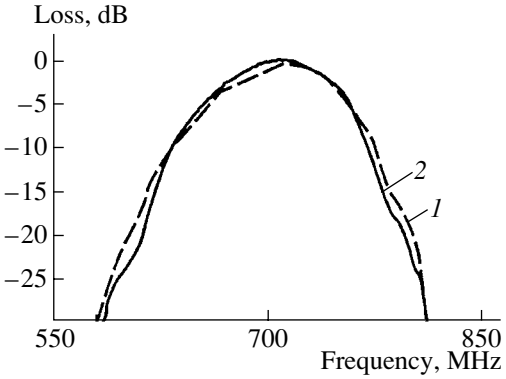


Fig. 8. (*1*) Experimental frequency responses of the pulse-compression dispersive delay line and (*2*) those calculated by the modified COM method.

the effect of spurious reflections on the frequency response of the DDLs, the thickness of the aluminum electrodes was as small as 0.075 μm. Also, the electrodes were placed in grooves etched in the $LiNbO_3$ substrate. Figure 7a shows experimental frequency responses of the pulse-forming DDL; those calculated by the modified COM method are plotted in Fig. 7b. The phase response $\Phi(f)$ in Fig. 7 is shown as the deviation from the given quadratic law: $\Delta\Phi = \Phi(f) - \pi G_0(f_0 - f)^2 - 2\pi T_0(f_0 - f)$, where $G_0 = 323$ MHz/μs is the slope of the dispersion curve and $T_0 = 0.9045$ μs is the filter delay at the frequency f_0.

The measured and calculated frequency responses of the pulse-compressing DDL are displayed in Fig. 8. In our calculations, the structure was divided into ≈400 channels, so that a further increase in the number did not change the results to within the graphical accuracy in Figs. 7 and 8. The calculations were performed on a 1-GHz Pentium-III PC. The calculated insertion losses were 18.2 dB; the measured value, 22 dB. The difference is apparently associated with the neglect of ohmic losses in the electrodes, which were only 0.075 μm thick.

CONCLUSION

The method developed in this paper can be used for designing SAW devices such as filters, dispersive delay lines, and resonators, on both "strong" piezoelectrics (lithium niobate and lithium tantalate) and "weak" piezoelectrics (quartz). In addition to the devices described, the method is applicable to designing SAW filters of any other type, including transversal filters, longitudinal-wave resonance filters, transverse-wave resonance filters, and SAW-resonator-based ladder-type filters.

REFERENCES

1. B. R. Potter and C. S. Hartmann, IEEE Trans. Sonics Ultrason. **SU-26**, 411 (1979).
2. V. F. Dmitriev and I. S. Mitrofanov, in *Proceedings of the International Symposium on Acoustoelectronics, Frequency Control and Signal Generation, Poland, 1998*, p. 463.
3. S. V. Birykov, G. Martin, V. G. Polevoi, *et al.*, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-42**, 612 (1995).
4. V. F. Dmitriev and B. A. Kalinikos, Radiotekh. Élektron. (Moscow) **33**, 2248 (1988).
5. V. F. Dmitriev and B. A. Kalinikos, Zh. Tekh. Fiz. **59** (1), 197 (1989) [Sov. Phys. Tech. Phys. **34**, 118 (1989)].
6. N. I. Muskhelishvili, *Singular Integral Equations* (Nauka, Moscow, 1968; Wolters-Noordhoff, Groningen, 1972).
7. S. G. Mikhlin, *Integral Equations* (OGIZ, Moscow, 1947).

Translated by A. Khzmalyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1166–1171. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 102–107.

ELECTRON AND ION BEAMS, ACCELERATORS

Spectral Properties of Interference Microwave Filters Based on Crossed Polarizing Gratings

A. V. Arzhannikov, S. A. Kuznetsov, and S. L. Sinitsky

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

e-mail: sakuzn@inp.nsk.su

Received January 23, 2002

Abstract—An analytic treatment and computer calculations of the spectral properties of novel interference microwave filters are carried out. A filter is represented by a set of plane-parallel polarizing gratings made of linear conductors whose orientation in neighboring gratings is chosen in a special way. The analysis is based on the assumption that the polarization properties of the gratings are perfect. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Polarizing gratings made of linear conductors are widely used in microwave devices operating in the mm- and sub-mm-ranges. In particular, such gratings are employed as mirrors in microwave Fabry–Perot interferometers (FPIs) [1–5].

In conventional microwave FPIs, the orientation of the conductors in mirrors–gratings is identical, which makes it possible to apply the theory of optical FPI. Specifically, the relative transmission bandwidth $\Delta\lambda/\lambda$ for a given interference order m is uniquely defined by the energy reflection coefficients R_1 and R_2 of FPI mirrors [6]:

$$\frac{\Delta\lambda}{\lambda} = \frac{1-R}{m\pi\sqrt{R}}, \qquad (1)$$

where $R = \sqrt{R_1 R_2}$.

Expression (1) shows that, for a given m, one can vary the relative bandwidth $\Delta\lambda/\lambda$ by using only mirrors with different reflectances. If a microwave FPI works as a filter with a variable transmission band, the replacement of mirrors–gratings causes significant difficulties. First, the necessity for system readjustment arises. Second, one must have a large number of gratings with different reflection coefficients.

To get rid of the above disadvantages, it is expedient to use a multibeam interferometer (interference filter) in which the conductors of neighboring plane-parallel polarizing gratings make an angle with each other. The advantage of interference filters with crossed gratings is the possibility of smoothly varying the spectral properties by merely varying the crossing angles without replacing the gratings.

In this work, we analyze the spectral properties of interference microwave filters with ideal polarizing gratings. The calculations are performed by the methods developed in [7].

STATEMENT OF THE PROBLEM

Consider the general case of an interference filter consisting of N arbitrarily crossed plane-parallel gratings made of linear conductors (Fig. 1). Let us number the gratings from left (first) to right (Nth). The XOY plane of the Cartesian coordinates XYZ coincides with the plane of the first grating, and the OZ axis is directed to the Nth grating. Let d_n be the spacing between the nth and $(n+1)$th gratings and $\alpha_n \in [0, \pi]$ be the counterclockwise-counted angle between the OX axis and the direction of the nth grating conductors. Since the properties of the filter depend on the relative crossing angles, we assume that the orientation angle α_1 of the first grating is constant and equals $\pi/2$.

We will also assume that the gratings are perfect polarizers. This means that the amplitude transmission and reflection coefficients for E- and H-polarized waves meet the equalities

$$\tau^E = 0, \quad \rho^E = -1, \quad \tau^H = 1, \quad \rho^H = 0. \qquad (2)$$

Physically, equalities (2) imply that the period of the conductors in the grating is negligibly small in comparison with the wavelength of incident radiation and that the conductivity of the conductor material is indefinitely high [8].

Let a plane monochromatic wave be incident from the left ($Z < 0$) on the filter orthogonally to the first grating. The amplitude of the electric field of the wave is E_0 and the absolute value of the wave vector is k. Hereafter, we assume that the incident wave is H-polarized relative to the first grating (Fig. 1). In this coordinate sys-

1063-7842/02/4709-1166$22.00

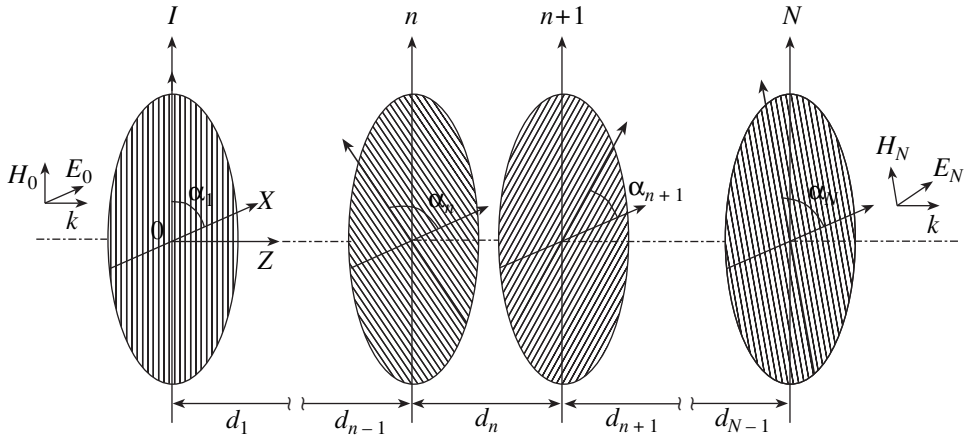


Fig. 1. Basic diagram of an interference microwave filter consisting of *N* arbitrarily crossed gratings.

tem, the Jones vector of the wave is

$$\mathbf{E}_0 = \begin{pmatrix} E_0 \\ 0 \end{pmatrix}.$$

Such a polarization of incident radiation complies with conditions (2). Note that the polarization considered is basically different from a conventional microwave FPI employing an *E*-polarized wave.[1] We are looking for the energy transmission coefficient of the filter. To neglect diffraction effects, we assume that the apertures of the gratings and the transverse size of the incident beam are much greater than the wavelength λ.

SPECTRAL PROPERTIES OF A FILTER WITH TWO CROSSED GRATINGS

The calculation of the transmission matrix $T_2^{\rightarrow}$ of a two-grating filter by the method of recurrence formulas [7] in view of equalities (2) yields

$$T_2^{\rightarrow} = \frac{(1-e^{2i\gamma})e^{i\gamma}\sin\alpha}{(1-e^{2i\gamma}\sin^2\alpha)}\begin{pmatrix} \sin\alpha & \vdots & 0 \\ \cdots & \cdots & \cdots \\ -\cos\alpha & \vdots & 0 \end{pmatrix},$$

where $\gamma = kd$, $d \equiv d_1$, and $\alpha \equiv \alpha_2$.

Recall that an arrow pointing rightward indicates that the *z* component of the wave vector of the incident wave is aligned with the *OZ* axis.

For an *H*-polarized wave, the energy transmission coefficient of the filter is found from the expression $\hat{T}_2^{\rightarrow} = |T_{2,11}^{\rightarrow}|^2 + |T_{2,21}^{\rightarrow}|^2$. The result is

$$\hat{T}_2^{\rightarrow} = \frac{4\sin^2\gamma}{\cos^2\alpha\cot^2\alpha + 4\sin^2\gamma}. \qquad (3)$$

Figure 2 plots the quantity $\hat{T}_2^{\rightarrow}$ versus the parameter γ/π for various orientation angles α of the conductors in the second grating.

The positions of maxima and minima in the curves $\hat{T}_2^{\rightarrow}(\gamma)$ are found from the conditions

$$\min: \gamma = n\pi, \text{ where } n = 0, 1, 2, \ldots;$$

$$\max: \gamma = n\pi + \pi/2.$$

The corresponding wavelengths are

$$\lambda_n^{\min} = \frac{2d}{n}, \quad \lambda_m^{\max} = \frac{2d}{n+1/2}.$$

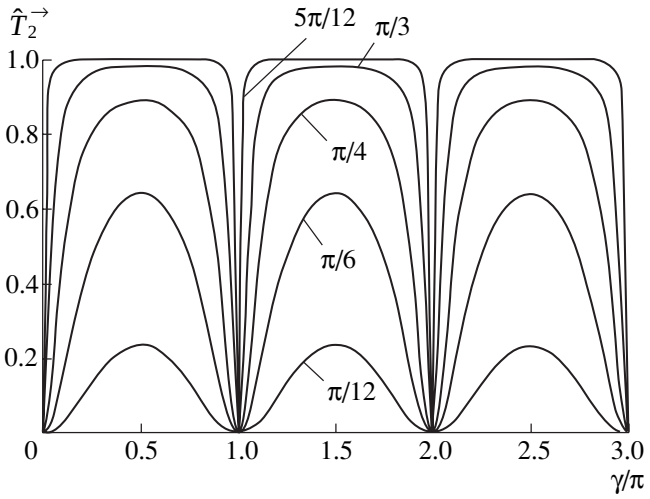


Fig. 2. Spectral transmission curves of a two-grating filter for various orientation angles α of the second grating. Numbers by the curves indicate the values of the angle ε, which is equal to α if $0 \le \alpha \le \pi/2$ or $(\pi - \alpha)$ if $\pi/2 \le \alpha \le \pi$.

[1] An *E*-polarized wave would be completely reflected from the first grating of the filter, so that other gratings would make no effect on its transmittance.

In minima, the filter transmission equals zero regardless of the value of α. In maxima, the transmission coefficient is $\hat{T}_{2\max}^{\rightarrow} = 4\sin^2\alpha/(1 + \sin^2\alpha)^2$. The function $\hat{T}_{2\max}^{\rightarrow}(\alpha)$ decreases monotonically from unity to zero with α varying from $\pi/2$ to 0 (or from $\pi/2$ to π).

Note wide interference maxima in the transmission curves at any value of α. Thus, we cannot design a narrow-band transmitting interference filter consisting of two ideal polarizing gratings. Nevertheless, an advantage of this structure is that it makes it possible to separate out narrow reflection bands of a desired width at α close to $\pi/2$ (i.e., at $\alpha_2 \cong \alpha_1$). In this case, the transmission coefficient of the filter is close to unity nearly throughout the spectral range of dispersion.

From (3), one can easily demonstrate that the relative width $\Delta\lambda/\lambda_n^{\min}$ of the reflection band depends quadratically on the grating misalignment δ when the misalignment is small ($|\delta| = |\pi/2 - \alpha| \ll 1$):

$$\frac{\Delta\lambda}{\lambda_n^{\min}} \cong \frac{p\delta^2}{n\pi}.$$

Here, the parameter p defines the reflection coefficient level used for calculating the bandwidth. Specifically, $p = 1$ for the half-maximum level and $p = 3$ for a level of 0.1.

SPECTRAL PROPERTIES OF A FILTER WITH THREE CROSSED GRATINGS

If conditions (2) are satisfied, the transmission matrix $T_3^{\rightarrow}$ of a three-grating filter is given by

$$T_3^{\rightarrow} = \frac{(1-e^{2i\gamma_1})(1-e^{2i\gamma_2})e^{i(\gamma_1+\gamma_2)}\sin(\alpha+\beta)\sin\alpha}{(1-e^{2i\gamma_2}\sin^2(\alpha+\beta))(1-e^{2i\gamma_1}\sin^2\alpha) - e^{2i(\gamma_1+\gamma_2)}\cos^2(\alpha+\beta)\cos^2\alpha}\begin{pmatrix} \cos\beta & \vdots & 0 \\ \cdots\cdots & & \\ -\sin\beta & \vdots & 0 \end{pmatrix},$$

where $\gamma_1 = kd_1$, $\gamma_2 = kd_2$, $\alpha \equiv \alpha_2$, and $\beta \equiv \pi/2 - \alpha_3$.

The calculation of the energy transmission coefficient of the filter by the expression $\hat{T}_3^{\rightarrow} = \left|T_{3.11}^{\rightarrow}\right|^2 + \left|T_{3.21}\right|^2$ yields

$$\hat{T}_3^{\rightarrow} = \frac{16\sin^2(\alpha+\beta)\sin^2\alpha\sin^2\gamma_1\sin^2\gamma_2}{\chi^2 + 4\sin^2\gamma_1\sin^2\gamma_2[\sin^2(\alpha+\beta) + \sin^2\alpha]^2}, \quad (4)$$

where

$$\chi = \sin(\gamma_1+\gamma_2)[\cos^2(\alpha+\beta) + \cos^2\alpha] + \sin(\gamma_1-\gamma_2)[\cos^2(\alpha+\beta) - \cos^2\alpha].$$

Consider two possible cases: $\beta = 0$ and $\beta \neq 0$.

(i) $\beta = 0$ (equal orientation of the outer gratings). Condition $\beta = 0$ makes it possible to reduce expression (4) to

$$\hat{T}_3^{\rightarrow} = \frac{4\sin^2\gamma_1\sin^2\gamma_2}{\cot^4\alpha\sin^2(\gamma_1+\gamma_2) + 4\sin^2\gamma_1\sin^2\gamma_2}. \quad (5)$$

Figure 3 shows typical $\hat{T}_3^{\rightarrow}$ versus $(\gamma_1 + \gamma_2)/\pi$ curves for various orientation angles α of conductors in the middle grating. It is seen that both the transmission and reflection bandwidths vary smoothly with the angle α. Narrow-band transmission is observed when the middle grating is almost orthogonal to the outer ones ($\alpha \ll 1$ or $(\pi - \alpha) \ll 1$). Narrow-band reflection occurs in the opposite case, when the angles of orientation of the three gratings are nearly equal ($|\pi/2 - \alpha| \ll 1$). Thus, a three-grating filter retains the advantage of a two-grating one. Note also that the minimum and maximum values of $\hat{T}_3^{\rightarrow}$ remain equal to zero and unity, respectively, for any value of α, which is the consequence of the ideality of gratings as polarizers.

It follows from expression (5) that the positions of minima and maxima in the transmission curve meet the conditions

$$\min: \gamma_j = n_j\pi, \text{ where } n_j = 0, 1, 2, \ldots; \quad j = 1, 2;$$
$$\max: \gamma_1 + \gamma_2 = m\pi, \text{ where } m = 1, 2, 3, \ldots; \quad m \neq n_j.$$

Then, the associated wavelengths are

$$\lambda_{n_j}^{\min} = \frac{2d_j}{n_j}, \quad j = 1, 2; \quad \lambda_m^{\max} = \frac{2(d_1+d_2)}{m}.$$

From expression (5), one can derive formulas for the relative bandwidth in the practically important narrow-band operating modes:

$$\frac{\Delta\lambda}{\lambda_{n_j}^{\min}} \cong \frac{p\delta^2}{n_j\pi} \quad (6)$$

for narrow-band reflection ($|\delta| \ll 1$, where $\delta = \pi/2 - \alpha$);

$$\frac{\Delta\lambda}{\lambda_m^{\max}} \cong \frac{4p\phi_m\varepsilon^2}{m\pi}, \quad (7)$$

where

$$\phi_m = \begin{cases} \cos^2\Delta_m, & m \text{ is odd} \\ \sin^2\Delta_m, & m \text{ is even,} \end{cases}$$

$$\Delta_m = \frac{\pi m}{2}\frac{(d_1 - d_2)}{(d_1 + d_2)},$$

for narrow-band transmission ($\varepsilon \ll 1$, where $\varepsilon = \alpha$ or $\varepsilon = \pi - \alpha$).

The parameter p in formulas (6) and (7) equals 1 and 3 for reflection (transmission) coefficient levels of 0.5 and 0.1, respectively.

At a given order of interference m, the relative bandwidth $\Delta\lambda/\lambda_m^{\max}$ can be varied by varying not only the crossing angle of the gratings but also the spacing between them owing to the presence of the phase parameter ϕ_m in formula (7). Therefore, the ratio $\Delta\lambda/\lambda_m^{\max}$ can vanish even at $\varepsilon \neq 0$. It is seen that the condition $\phi_m = 0$ is satisfied when minima with different subscripts j coincide. The resonance values of the spacing (for which $\phi_m = 0$) are related to the wavelength of a maximum as

$$d_1 = \frac{m \pm l}{4}\lambda_m^{\max}, \quad d_2 = \frac{m \mp l}{4}\lambda_m^{\max}.$$

Here, l is a positive integer of the same parity as m. It varies from 1 to m or from 0 to m if m is odd or even, respectively.

In practice, the free-dispersion spectral range of an interference filter should be maximized. For this reason, it is expedient to take the same distances d_1 and d_2 so as to eliminate even maxima.

(ii) $\beta \neq 0$. In the case of an arbitrary misalignment β of conductors in the outer gratings, the spectral curve of a three-grating filter transmittance is intermediate between the above two variants $\beta = 0$ and $\beta = \pi/2 - \alpha$ (the latter corresponds to a two-grating filter). Hence, the analysis of the case $\beta \neq 0$ does not provide fundamentally new solutions. The only exception is that narrow-band reflection can take place at any value of α if the orientations of the second and third gratings are close ($|\pi/2 - (\alpha + \beta)| \ll 1$). For this reason, we restrict our consideration to some remarks regarding the optimum value of β.

As follows from formula (4), if $\beta \neq 0$, the filter has the maximum transmission at wavelengths meeting the condition $\chi = 0$. The maximum transmission coefficient is given by

$$\hat{T}_{3\max}^{\rightarrow} = \frac{4\sin^2(\alpha + \beta)\sin^2\alpha}{[\sin^2(\alpha + \beta) + \sin^2\alpha]^2}. \quad (8)$$

Expression (8) shows that in the case of narrow-band transmission ($\varepsilon \longrightarrow 0$) and a nonzero value of the angle β, the maximum transmission coefficient falls to zero as $T_{3\max}^{\rightarrow}\big|_{\varepsilon \to 0} \cong 4\varepsilon^2/\sin^2\beta$. This sets a fundamental limitation on the ultimate transmission bandwidth achievable by varying α. Hence, the case $\beta = 0$, for which, as was demonstrated above, $\hat{T}_{3\max}^{\rightarrow} = 1$ at any value of α, is optimal.

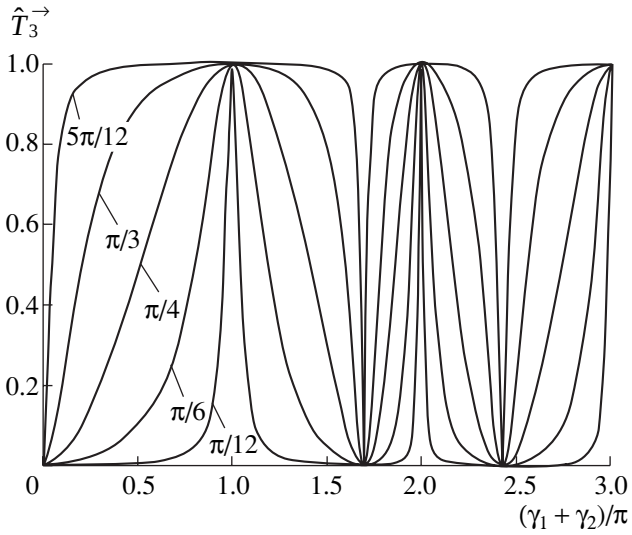


Fig. 3. Spectral transmission curves of a three-grating filter for various orientation angles α of the middle grating for $\beta = 0$ and $d_1/d_2 = 0.7$. Numbers at the curves indicate the values of the angle ε, which is equal to α if $0 \le \alpha \le \pi/2$ or $(\pi - \alpha)$ if $\pi/2 \le \alpha \le \pi$.

Calculations of the relative transmission bandwidth favor the view that the same orientation of the outer gratings is the optimal condition. If ε and $|\beta| \ll 1$, the relative bandwidth takes the form (cf. (7))

$$\frac{\Delta\lambda}{\lambda_m^{\max}} \cong \frac{2p\phi_m[\varepsilon^2 + (\varepsilon + \beta)^2]}{m\pi}. \quad (9)$$

From (9), at a given ε, the bandwidth is minimum at $\beta = 0$. Figure 4 shows curves $\hat{T}_{3\max}^{\rightarrow}(\varepsilon)$ and plots of spectral resolution for various values of β.

SPECTRAL PROPERTIES OF COMPLEX FILTERS

An increase in the number of gratings in an interference filter complicates the interference pattern in the structure and gives rise to new features in its transmission (reflection) spectrum. For this reason, an interference filter containing more than three gratings will be called complex.

In addition to the possibility of controlling the shape of a transmission (reflection) curve, complex filters offer such an important advantage as much steeper transmission (reflection) band edges in comparison to two- and three-grating filters. Let us demonstrate this with a five-grating filter.

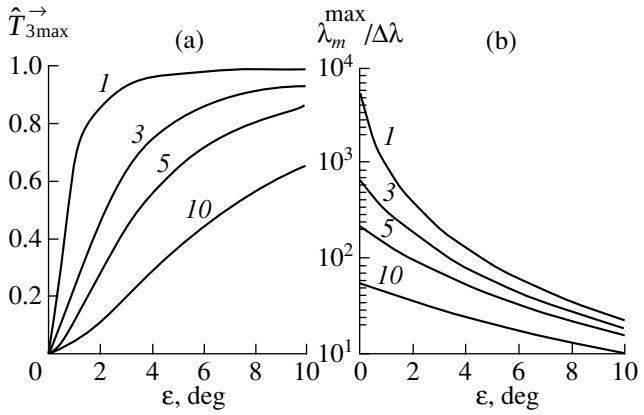


Fig. 4. (a) Maximum energy transmission coefficient and (b) spectral resolution of a three-grating filter versus the angle ε. Numbers by the curves indicate the values of the angle β (in degrees). $m = 1$, $d_1/d_2 = 1$.

We represent a five-grating filter as a combination of two optimally configured three-grating filters and start with the case when $d_1 = d_2 = d_3 = d_4 \equiv d$, $\alpha_1 = \alpha_3 = \alpha_5 = \pi/2$, and α_2 and α_4 are free parameters. It can be shown that the transmission spectrum of such a filter has principal maxima corresponding to the condition $\gamma = m\pi$ ($\gamma = 4\pi d/\lambda$, $m = 1, 3, 5,\ldots$) and extra ones lying symmetrically on either side of the principal peaks. The positions and amplitudes of the extra maxima depend on the orientation angles of the second and fourth gratings. In the most interesting case $\alpha_2 = \alpha_4 \equiv \alpha$, the amplitude of the extra maxima equals unity and their steepness is maximum (Fig. 5). For α ranging between π/4 and 3π/4, the transmission coefficient of the filter in valleys between the principal and extra maxima (hereafter denoted as $\hat{T}_{5\min}^{\rightarrow}$) is no less than 0.86. This makes it possible to use a five-grating filter as a band-pass filter with a transmission bandwidth approximately equal to the spectral interval ΔΓ between the extra maxima.

A comparison of the transmission curves for three- and five-grating filters shows that the latter has an apparent advantage when operating in the wide-band mode ($\varepsilon \approx 1$), since the three-grating filter has transmission bands with smooth edges (Fig. 5). In the narrow-band mode ($\varepsilon \longrightarrow 0$), the five-grating filter is still superior in terms of the steepness of the edges. However, in this case, the value of $\hat{T}_{5\min}^{\rightarrow}$ decreases to zero and the complex filter selects three closely spaced bands instead of one. This unwanted effect can be eliminated by changing the operating conditions for the five-grating filter.

The analysis shows that one can satisfy the requirement $\hat{T}_{5\min}^{\rightarrow} \cong 1$ and vary the value of ΔΓ by changing the orientation angle α_3 of the third grating. The value of angle α must provide reasonable values of $\hat{T}_{5\min}^{\rightarrow}$ and steepness of the band edges. Consider the following configuration of a five-grating filter: $d_1 = d_2 = d_3 = d_4 \equiv d$, $\alpha_1 = \alpha_5 = \pi/2$, $\alpha_2 = \alpha_4 \equiv \alpha$, and $\alpha_3 \equiv \pi/2 - \beta$. The energy transmission coefficient in this case is given by

$$\hat{T}_5^{\rightarrow} = \hat{T}_5^{\leftarrow} = \left| \frac{(1 - e^{i\gamma})\sin^2\alpha\sin^2(\alpha + \beta)}{[1 + e^{i\gamma}\cos^2\alpha][1 + e^{i\gamma}(1 + e^{i\gamma})\cos^2\alpha + e^{i\gamma}\cos 2(\alpha + \beta)]} \right|^2, \quad (10)$$

where $\gamma = 4\pi d/\lambda$.

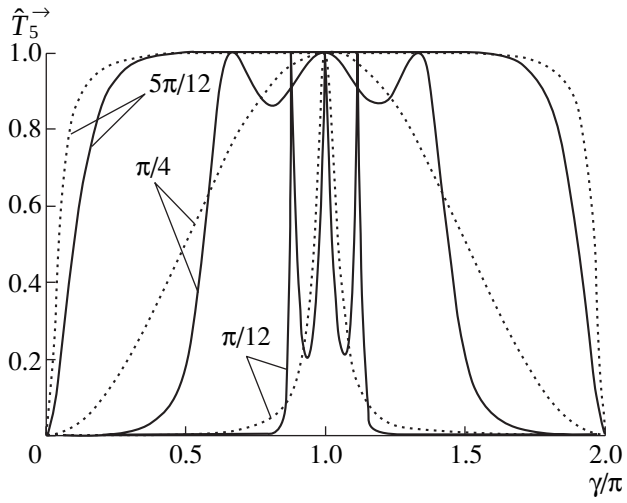


Fig. 5. Spectral transmission curves of five-grating (solid lines) and three-grating (dashed lines) filters for various values of the angle ε (numbers by the curves).

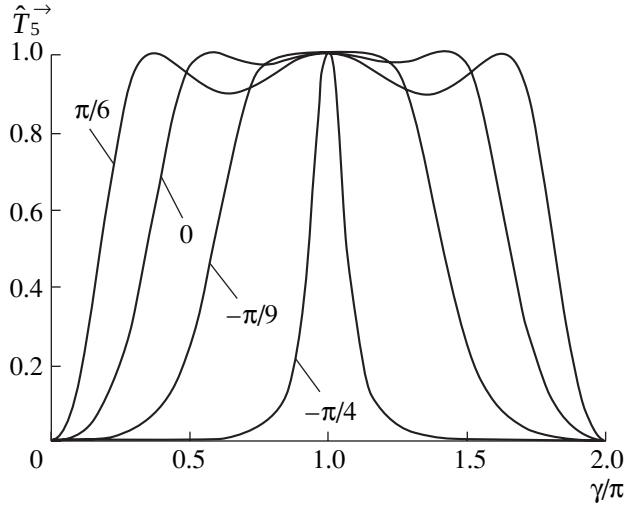


Fig. 6. Variation of the spectral transmission curves of a five-grating filter with the orientation of the third grating for $\alpha = \pi/6$. Numbers by the curves indicate the values of β.

Figure 6 shows the spectral curves of the filter transmission for several values of the angle β.

In the wide-band mode ($|\phi| \approx 1$, where $\phi = \alpha + \beta$ or $\phi = \pi - (\alpha + \beta)$), the spectral interval $\Delta\Gamma$ between the extra maxima makes a major contribution to the bandwidth. According to expression (10), the value of $\Delta\Gamma$ expressed in units of γ is

$$\Delta\Gamma = 2\left(\pi - 2\arccos\frac{\sqrt{D}}{2|\cos\alpha|}\right),$$

where $D = 2\sin^2(\alpha + \beta) - \sin^2\alpha[\sin^2(\alpha + \beta) + \sin^2\alpha]$.

The value of $\Delta\Gamma$ is maximal when the second (fourth) and third gratings are equally oriented ($\alpha + \beta = \pi/2$). If the third and second (fourth) gratings are crossed, the value of $\Delta\Gamma$ decreases and the principal and extra maxima merge together at $D = 0$. The narrow-band mode arises when $|\phi| \ll 1$ (the crossing angle between the second (fourth) and third gratings is close to $\pi/2$). In this mode, $D < 0$ and the relative transmission bandwidth has the form

$$\frac{\Delta\lambda}{\lambda_m^{\max}} \cong \frac{4p}{m\pi}\left(\frac{\phi}{\sin\alpha}\right)^2.$$

CONCLUSION

We theoretically studied the spectral properties of interference filters consisting of several crossed polarizing gratings. The analysis made it possible to obtain the dependences of the transmission (reflection) band shape and width on the crossing angles between the gratings, which opens up possibilities for filter applications.

It was demonstrated that a simple two-grating filter can operate as narrow-band only in the reflection node. In its turn, a three-grating filter can be narrow-band in both the reflection and transmission modes. More complex filters, retaining the advantages of three-grating filters, provide an additional possibility to vary both the shape and the steepness of the transmission (reflection) spectral curves. In this work, this property was demonstrated with a five-grating filter.

The analysis was carried out under the assumption that the gratings operate as perfect polarizers and that the angular divergence of radiation is absent. In this approximation, the ultimate spectral resolution $(\lambda/\Delta\lambda)_{\max}$ attained at the optimum mutual orientation of the gratings is limited only by diffraction effects due to a finite transverse size of the gratings. In addition, the value of $(\lambda/\Delta\lambda)_{\max}$ is bounded above when the gratings are misaligned with respect to the optimum orientation. For example, a misalignment of 1° between the outer gratings of a three-grating filter gives $(\lambda/\Delta\lambda)_{\max} \cong 10^3$ for a maximum transmission coefficient of about 70% (Fig. 4).

In the next work, the effect of imperfect polarizing gratings and of the angular divergence of incident radiation on the spectral characteristics of the filters will be considered.

REFERENCES

1. F. A. Korolev and V. A. Gridnev, Opt. Spektrosk. **16**, 335 (1964).
2. A. V. Chernetskiĭ, O. A. Zinov'ev, and O. V. Kozlov, *Equipment and Methods for Plasma Studies* (Atomizdat, Moscow, 1965).
3. E. A. Vinogradov, E. M. Dianov, and N. A. Irisova, Pis'ma Zh. Éksp. Teor. Fiz. **2**, 323 (1965) [JETP Lett. **2**, 205 (1965)].
4. V. Ya. Balakhanov, V. D. Rusanov, and A. R. Striganov, Zh. Tekh. Fiz. **35**, 127 (1965) [Sov. Phys. Tech. Phys. **10**, 96 (1965)].
5. A. A. Bagdasarov, V. V. Buzankin, N. L. Vasin, *et al.*, in *Diagnostics of Plasma: Collection of Articles* (Énergoizdat, Moscow, 1981), Vol. 4 (1), pp. 141–146.
6. G. V. Rozenberg, *Thin-Film Optics* (GIFML, Moscow, 1958).
7. A. V. Arzhannikov and S. A. Kuznetsov, Zh. Tekh. Fiz. **71** (12), 1 (2001) [Tech. Phys. **46**, 1489 (2001)].
8. V. P. Shestopalov, L. P. Litvinenko, S. A. Masalov, and V. G. Sologub, *Wave Diffraction by Gratings* (Khar'kovskiĭ Univ., Kharkov, 1973).

Translated by A. Chikishev

Technical Physics, Vol. 47, No. 9, 2002, pp. 1172–1175. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 108–111.

**SURFACES,
ELECTRON AND ION EMISSION**

Energy Distribution of Field Emission Electrons from Carbon Nanocrystals

V. M. Lobanov[1] and Yu. M. Yumaguzin[2]

[1] *Bashkir State Agrarian University, Ufa, 450001 Russia*

[2] *Bashkir State University, Ufa, 450074 Russia*

e-mail: bgau@soros.bashedu.ru

Received May 10, 2001; in final form, January 14, 2002

Abstract—Using a field electron microscope and a field electron dispersion energy analyzer, carbon nanocrystals contained in polyacrylonitrile (PAN) fibers are studied. Transition of the emitting nanocrystals into the second stable state has been discovered, corresponding to an emission current and field electron energy distribution of magnitudes lower by nearly an order of magnitude, with an additional low-energy peak. Heating the samples at 750°C restores the initial characteristics of the field electron energy distribution and current–voltage characteristics of the nanocrystals. The forbidden bandwidth of the nanocrystals has been determined. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The emergence in the 1970s of technologies for production of thin carbon fibers stimulated an extensive research effort aimed at creating a carbon field emission cathode featuring a high melting point, low adsorption of residual gases in vacuum, high conductivity, and mechanical strength. However, the attendant problems still have not been overcome: (1) forming a tip with a smooth surface and regular image; (2) energy analysis of the field emission electrons have revealed two kinds of energy distributions: normal distribution, explainable in terms of the model of free electrons and a broad distribution with two maxima. Energy analyses of the field electron emission were complicated by the heterogeneity of the graphitized carbon electrodes, where both crystalline and amorphous phases of carbon were present. The form of energy distribution could not be related to the phase composition of the emitter spot probed nor could the energy band structure of this region be accounted for. It is known that nanocrystals of graphitized carbon materials have a bandwidth that decreases from 0.5–0.3 to <0.01 eV as the thermal treatment temperature is raised from 600 to 2300°C, and the nanocrystal size increases from ~1.5 to ~15 nm [1].

In this work, the object of study was PAN carbon fibers (carbon fiber prepared from a synthetic polyacrylonitrile fiber) annealed at a temperature of 900°C. The base planes of the nanocrystals it contains are oriented along the fiber axis; the size of nanocrystals was 1.5–8 nm; and the forbidden band width was in the range 0.15–0.03 eV [1, 2].

EXPERIMENTAL

The studied fiber of length ~1 mm was fixed with an aquadag to a narrow strip of tantalum foil welded to a tungsten arc and, after drying, was heated in vacuum at a temperature of 750°C measured with an optical pyrometer. The field cathode obtained was introduced via a locking unit into a USU-4 superhigh-vacuum chamber equipped with a Muller electron-microscope projector and an electrostatic energy dispersion analyzer whose secondary-electron multiplier was operated in the pulse-count regime.

After locking, "forming" was carried out under rough vacuum conditions by raising the total emission current to ~150 μA, whereby nanocrystals could be isolated on the fiber emitting surface due to the sputtering of the amorphous carbon by ions of the residual gas [3]. The picture observed on the microscope monitor gave evidence of intense electron emission from the entire carbon fiber end face. A separate bright spot was projected onto the energy analyzer probing orifice. Under a superhigh vacuum, the sample emission surface was additionally cleaned by heating the tungsten arc to 750°C (the surface of carbon materials is almost completely cleaned by heating to a temperature of ≈650°C [4]). Measurements of the energy distribution of field emission electrons (EDFEE) were performed using a technique described in [5], and the Fermi level position was determined using a reference tungsten emitter. Current–voltage characteristics (IVC) of the probe current were plotted as values of the electron current passing the energy analyzer and corresponding values of the anode voltages U_a.

1063-7842/02/4709-1172$22.00 © 2002 MAIK "Nauka/Interperiodica"

EXPERIMENTAL RESULTS

After heating the emitter in a superhigh vacuum, the probe current became stable enough for EDFEE measurements. Figure 1a shows how EDFEE varied with anode voltage: the energy distribution shifted in increments proportional to steps of U_a. This shift of EDFEE to the region of low energies was accompanied by considerable widening of its high-energy edge. After scanning the EDFEE at U_a = 3980 V, the brightness of the emission spot dropped drastically and repeated scanning of the EDFEE revealed an additional low-energy peak lower in energy than the main one by ~0.45 eV.

In the course of the study, one more stage was added to the sample preparation procedure after forming in a rough vacuum: under a superhigh vacuum, the field emission cathode was repeatedly stripped of its surface layer by applying current pulses of ~130 μA until a uniform stable bright spot was obtained in the central area of the emission image. This spot was then placed opposite the energy analyzer probing orifice.

After heating in a superhigh vacuum, the EDFEE of this second type of sample had one peak and shifted to the low-energy region with increasing U_a. As in the first case, considerable broadening of the high-energy edge was observed, and at U_a = 3980 V the emission current and emission-spot brightness dropped sharply, and an additional, low-energy peak appeared in the field electron energy distribution at ~0.45 eV below the main peak. With decreasing U_a, the relative magnitude of the low-energy peak in the EDFEE dropped, and the energy distribution uniformly shifted to the high-energy region.

Heating the sample at 750°C restored the starting EDFEEs and IVCs of the probe current, but after U_a was raised to 3800 V (to 3700 V after the third heating), the probe current again dropped, and in the EDFEE an additional low-energy peak appeared at an energy lower than the main peak by ~0.45 eV. The measurements yielded pairs *1–2*, *3–4*, and *5–6* of the current–voltage characteristics of the probe current seen in Fig. 1b, corresponding to EDFEE distributions with one (I) and two (II) peaks (Fig. 1a).

The (I ⟶ II) transition took place as a particular value of the anode current, different in each sample, was reached. Simultaneously, an additional energy peak at 0.45 eV below the main one emerged in the EDFEE, and the IVC of the probe current shifted into a lower position. Heating the samples at 750°C returned the EDFEEs and IVCs to their initial states and caused the (II ⟶ I) transition.

To check the correspondence between the increments in the shift of the energy distribution of field emission electrons and the steps of the anode voltage U_a, separate measurements of the EDFEE were carried out after prolonged heating of the sample with a readjusted secondary electron multiplier in the energy analyzer. The results of these measurements are shown in Fig. 2.

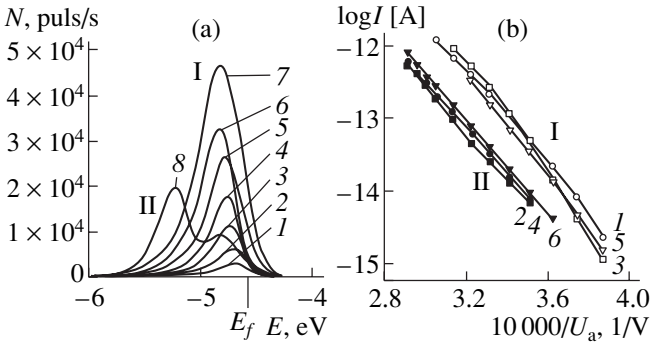


Fig. 1. Increase in U_a shifted EDFEE to lower energies; at U_a = 3980 an EDFEE I ⟶ IEDFEE II transition occurred (a); and cr-voltage characteristic IVC I and IVC II of a nanocrystal corresponding to EDFEE I and IEDFEE II (b). a—U_a, V: 3620 (*1*); 3680 (*2*); 3740 (*3*); 3800 (*4*); 3860 (*5*); 3920 (*6*); 3980 (*7*, *8*).

RESULTS AND DISCUSSION

The results obtained point to an interrelationship between the energy distributions of the field emission electrons from graphitized carbon cathodes with one and two peaks.

EDFEE I was observed in samples heated at T = 750°C in the absence of electric field. EDFEE II emerged at a current corresponding to some threshold value of the anode voltage. The EDFEE I ⟶ EDFEE II transition was accompanied by a drop in the probe current by nearly one order of magnitude. The anode current drop and the EDFEE I ⟶ EDFEE II transition were preceded by a shift of the energy distribution to the low-energy region by steps proportional to increments of anode voltage U_a. This kind of EDFEE shift in the case of semiconductor emitters is evidence of the Zener effect [6].

For appreciable electron tunneling from the valence band into the conduction band in a semiconductor with a forbidden bandwidth of $\Delta E_g \approx 1$ eV, the required field strength is $\sim 3 \times 10^7$ V/m. In our experiment, the voltage drop across the emitting carbon nanocrystal measured as a shift of the top of EDFEE (Fig. 3a) at low U_a, where the shift is not yet linear, reached ~150 mV, producing in PAN fiber nanocrystals of 1.5–8.0 nm in size an electric field of 10^7–10^8 V/m, which is quite sufficient for electron tunneling from the valence band into the conduction band.

It is know that in semiconductor emitters, the voltage drop across an emitter corresponding to the linear region of the IVC in Fowler–Nordheim coordinates (low anode voltage) is approximately equal to the forbidden bandwidth. Estimates of the voltage drop in the samples studied by the nonlinear portions of shift plots (Fig. 3a), corresponding to linear portions of IVC

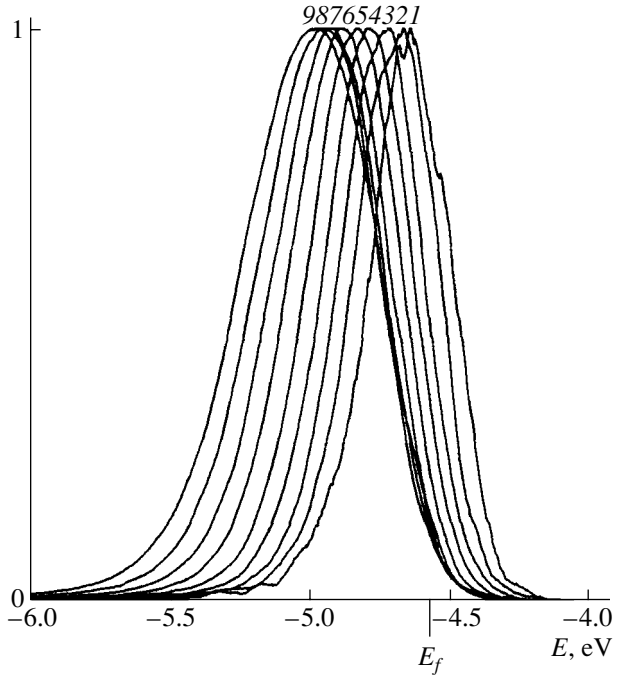


Fig. 2. The incremental shift of the top and low-energy edge of EDFEE from a carbon nanocrystal is proportional to the anode voltage step. U_a, V: 2800 (*1*); 2900 (*2*); 3000 (*3*); 3100 (*4*); 3200 (*5*); 3300 (*6*); 3400 (*7*); 3500 (*8*); 3600 (*9*).

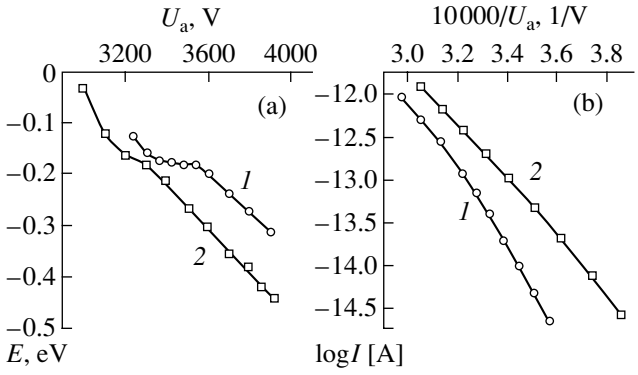


Fig. 3. Shift of the top of EDFEE I of samples *1* and *2* relative to the Fermi level (a) and IVCs of the probe current of samples *1* and *2* corresponding to EDFEE I (b).

(Fig. 3b), give values of 70 and 150 meV for the lower limits of the forbidden bandwidth of the emitting nanocrystals in the samples studied.

In the studies of energy distributions of field emission electrons emitted from semiconductors, the main problem in obtaining an EDFEE with two peaks corresponding to the bottom of the conduction band and the top of the valence band was interference from the surface states of electrons localized within the forbidden band [7]. The contribution to the EDFEE from these surface states was so high that it concealed the true energy distributions of electrons in a semiconductor, so that the energy distributions of field emission electrons with two peaks corresponding to the bottom of the conduction band and the top of the valence band could not be reliably determined.

Figure 4a shows the way in which the shape of the top of EDFEE in the samples studied changed at the lowest anode voltages. Similar changes of the top of EDFEE with anode voltage were also observed in other samples displaying at large anode voltages a linear shift of the energy distribution to the low-energy region of EDFEE with a broad top (Fig. 4a) preceding the Zener tunneling of electrons. We relate these changes to the potential barrier height value at which the top of EDFEE is formed by electrons tunneling both from the bottom of the conduction band and the top of the valence band, as well as, regrettably, by a rather intense flow of electrons emitted from the surface states localized within the forbidden band, for which reason the forbidden band cannot be resolved in the EDFEE. An estimate of the upper limit of the forbidden band of emitting nanocrystals in the samples studied from the width of medium EDFEEs gives values of 100 and 150 meV, respectively.

Thus, the forbidden bandwidth of the emitting nanocrystals in type-1 samples was found to be ΔE_g = 70–100 meV, and in type-2 samples, $\Delta E_g \approx 150$ meV, in fair agreement with the range of ΔE_g values of graphitized carbon nanocrystals annealed at 900°C.

It should be noted that throughout this study the linear shift of EDFEE into the low-energy range with U_a, the presence of the Zener effect, was considered as an indication that the probed spot at the emitting surface belonged to a nanocrystal and not to the amorphous component of the emitter.

Common features of EDFEE I behavior should also be noted: apart from the observed shift of the energy distribution to the low-energy region in steps proportional to the increment in U_a, at the lowest U_a values, the low-energy edge of EDFEE narrowed, and the energy distribution became more symmetrical. With increasing U_a, the linear shift of the top of EDFEE was accompanied by a linear shift of its low-energy edge; the shift of the high-energy edge to lower energies at large U_a values ceased, and its broadening towards the high-energy region began (Fig. 2). This sort of broadening of the EDFEE high-energy edge usually occurred as a result of the heating of the tungsten arc by an electric current; therefore, in this case broadening of the EDFEE high-energy edge was accounted for as the heating of the electrons and lattice by the penetrating electric field. In addition, at the highest U_a the broadening rate of the high-energy edge increased considerably. The broadening process of EDFEE I terminated with a drastic drop in the probe current and the emergence of an additional low-energy peak, i.e., with the EDFEE I $\longrightarrow$ EDFEE II transition (Fig. 1a). The good coincidence of the superposed high-energy edges of EDFEE I and EDFEE II measured at an anode voltage of 3980 V, corresponding to the I $\longrightarrow$ II transition

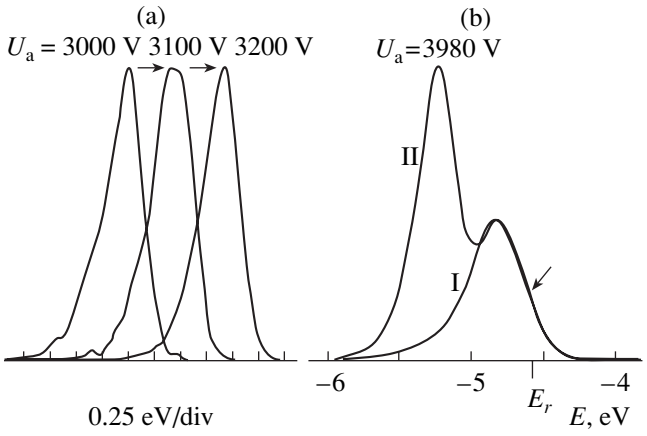


Fig. 4. (a) Change of form of the top of EDFEE I at lowest U_a values, and (b) coincidence of the high-energy edges of EDFEE I and EDFEE II measured at U_a for the I ⟶ II transition.

(Fig. 4b), showed that the broadening of the high-energy edge of EDFEE I held out after the transition of the carbon nanocrystals into the second stable state.

Taking into account the above, the following interpretation of the observed change of the energy distribution of field electrons emitted from a carbon nanocrystal is proposed.

(1) At low anode voltages, the probe current is comprised mainly of electrons tunneling from the valence band and the surface electron states within the forbidden band of the nanocrystal. An increase in voltage and the penetration of the electric field into the near-surface region of the nanocrystal cause band bending and Zener tunneling of the electrons from the valence band into the conduction band, increasing in the probe current the fraction of electrons tunneling from the conduction band and making the low-energy edge narrower and the EDFEE symmetry higher.

(2) A voltage drop across the emitting nanocrystal causes its heating by Joule heat and thermal-field reconstruction, producing at its emitting surface two groups of surface electron states near the Fermi level, one of which, located in the forbidden band and the conduction band, impedes electron emission and reduces the height of the main EDFEE peak, and the other, localized in the valence band, produces an additional low-energy peak 0.45 eV below the main peak. After thermal-field reconstruction of the emitting carbon nanocrystal, the fraction of electrons tunneling from the conduction band in the emission current from the nanocrystal decreased.

CONCLUSION

In the study of a carbon nanocrystal using an energy dispersion analyzer of field emission electrons, corresponding EDFEEs and IVCs of nanocrystals were compared. In a carbon nanocrystal with basal planes oriented in the direction of emission, two stable states of the field electron emission were observed whose emission currents differed by almost an order of magnitude; corresponding EDFEEs with one or two maxima were also observed. The transition of the nanocrystal from the first stable state into the second, accompanied by a decrease in the emission current and appearance in the EDFEE of an additional low-energy peak ~0.45 eV below the main peak, was reproduced in detail in several samples. Heating at 750°C restored the initial EDFEEs and IVCs of nanocrystals.

The transition of the emitting nanocrystal from the first stable state into the second was preceded by the Zener effect, which is evidence of a narrow forbidden band; its width was estimated for two samples: 70–100 meV for the one and ~150 meV for the other.

REFERENCES

1. S. V. Shulepov, *Physics of Graphitized Carbon Materials* (Metallurgiya, Moscow, 1972).
2. A. S. Fialkov, *Graphitized Carbon Materials* (Énergiya, Moscow, 1979).
3. B. V. Bondarenko, E. S. Bakanova, A. Yu. Cherepanov, and E. P. Sheshin, Radiotekh. Élektron. (Moscow) **30**, 2234 (1985).
4. Ph. L. Walker, L. G. Austin, and J. J. Tietjen, in *Chemistry and Physics of Carbon*, Ed. by Ph. L. Walker (Marcel Dekker, New York, 1965; Mir, Moscow, 1969).
5. R. Z. Bakhtizin, V. M. Lobanov, and Yu. M. Yumaguzin, Prib. Tekh. Éksp., No. 4, 247 (1987).
6. I. L. Sokol'skaya and G. P. Shcherbakov, Fiz. Tverd. Tela (Leningrad) **3**, 167 (1961) [Sov. Phys. Solid State **3**, 120 (1961)].
7. W. B. Shepherd and W. T. Peria, Surf. Sci. **38**, 461 (1973).

Translated by B. Kalinin

Technical Physics, Vol. 47, No. 9, 2002, pp. 1176–1181. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 112–118.

SURFACES, ELECTRON AND ION EMISSION

Impurity Composition and Decontamination of the Cylindrical Surface of CVD Polycrystalline Tungsten

A. V. Bravarets, B. M. Zykov, V. N. Zykova, V. N. Lebedev, and Yu. K. Udovichenko

Sukhumi Physicotechnical Institute, Abkhazian Academy of Sciences, Sukhumi, Abkhazia, 384914 Georgia

Received September 18, 2001

Abstract—The impurity composition on and near the cylindrical surface of a CVD polycrystalline tungsten film covering Nb + 1% Zr alloy is determined with SIMS and AES methods. The alloy is used as the collector material in thermionic thermal-to-electrical energy converters of space-borne nuclear power plants. A comparison with the impurity composition of a reference plane-parallel relatively perfect W(110) single crystal is made. At all stages of milling the surface by He^+ and Ar^+ ion beams, the coating is contaminated greater than the reference. Even after the first heating *in vacuo* to 1625 K, all the impurities, except carbon, oxygen, and niobium, are removed. The three impurities named are not removed by heating up to 2075 K. Niobium is shown to diffuse toward the surface through a W coating of thickness up to 30 μm. At temperatures above 1925 K, the material is heavily sputtered. After 50 oxidation cycles, the carbon content in the coating is greatly reduced. However, the subsequent 25 cycles fail in further decreasing the carbon content. The fact that oxygen cannot be removed from the surface by heating to 1925–2075 K suggests the presence of high-temperature surface Nb_2W_2O suboxide. It is recommended that the purity of the W coating raised in order to improve its adhesion to the O–Cs film. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The elemental composition and the chemical state of the surface of the electrodes in thermionic thermal-to-electrical energy converters (TECs) have a pronounced effect on their performance. Oxygen is known as the only nonmetallic impurity on the surface of the electrodes (especially, collector) in TECs with a cesium filler [1]. In general, the most harmful surface impurities in emission electronics are carbon and sulfur [2]. For example, the presence of MoC or Mo_2C carbides on the Mo(110) surface increases the work function at least by 0.9 eV, while the presence of CO raises this parameter by 1.63 eV [2]. If sulfur is present only on the surface and the material can be heated to 1650–1850 K, this element can be removed merely by heating *in vacuo*. Carbon, however, cannot be removed in this way from transition metals (Cr, Mo, W, Re, etc.), which do not dissolve oxygen in the lattice. This is because the sublimation temperature of carbides, which form when carbon combines with the substrate material, exceeds the melting point of these metals. Therefore, to remove carbon from near-surface layers, additional procedures such as oxidation, heating in hydrogen, or milling by inert gas ions (usually Ar^+) are applied. However, even they cannot provide complete cleaning of the surface from C if it is dissolved in the bulk of the material. Upon heating, volume carbon diffuses toward the surface and segregates on it in concentrations much higher than volume concentrations. To avoid such a situation, the electrodes must be made of extra-pure materials and pumping facilities operating with hydrocarbons must be excluded.

At present, a TEC collector is usually made of Nb + 1% Zr alloy covered by a 30- to 40-μm-thick polycrystalline W film. Therefore, the determination of impurities near and on the electrode surface, as well as the development of decontamination methods, would allow researchers to refine the electrode fabrication technology and thereby to improve the TEC efficiency.

EXPERIMENTAL METHODS

Three specimens measuring 10 × 10 mm were cut by a diamond tool from the (visually) most perfect part of the cylindrical collector in compliance with vacuum hygiene requirements. Two of them and the reference (a plane-parallel W(110) single crystal) were placed in an MS 7201 M secondary ion mass spectrometer. The reference identical to that used in [3] was made according to the requirements for specimens to be investigated by LEED. That is, the single-crystalline material has a small concentration of defects and impurities in the volume and the misorientation between the working surface, and the face selected was no more than 30'. In addition, layers damaged by mechanical grinding and polishing were removed by electropolishing. The impurity content in the reference at depths of 15–20 μm found previously by the SIMS method in an IMS-3F CAMECA instrument was the following (in ppm): H, 20; C, 50; O, 20; Na, 300; Al, 0.7; Si, 3.7; K, 100; Ca, 0.6; Ti, 0.2; Ni, 1.5; Cu, 0.8; Mo, 0.01; and Ta, 3. Before

1063-7842/02/4709-1176$22.00

being placed into the instrument, the specimens were rinsed in methylene chloride to remove fats and in boiling distilled water to remove salts. The specimens in the MS 7201 M instruments were not heated. The mass spectrometer was first adjusted to another material, usually molybdenum (six specimens can be placed in the instrument). Then the specimen to be studied was placed under the He^+ ion gun and the mass spectrum was recorded at once. In this way, the surface milling by the ion beam was minimized when initial contaminations were detected.

Under He^+ and Ar^+ ion bombardment, the impurity distribution over the specimen surface was also determined. For this purpose, we moved the probing ion beam over the surface by varying the specimen potential and the potential of the focusing electrode of the gun without changing the specimen position in the holder. The maximal displacement of the ion spot on the specimen was half the beam diameter (≈2 mm). Simultaneously, the instrument was adjusted to the peak intensity of an impurity selected. In the spectrum thus obtained, the intensities of this and other impurities were compared. Of special interest was the relative intensity of W.

During recording of the spectra, the residual gas pressure was no more than 5×10^{-6} Pa and that of the working gas, 12×10^{-2} Pa. Pumping facilities were free of hydrocarbons. The ion current beam was 5–20 μA; the ion spot area on the specimen, ≈0.1 cm^2; the He^+ or Ar^+ ion energy, 4 keV; and the angle of incidence of the beam on the specimen, 50°. As working gases, extrapure-grade Ar and He were employed. The total exposure time of the specimens including the spectrum recording time was usually about 20 min. In the case of Ar^+ ions, however, this time might be as long as 4 h. The spectra were recorded first with He^+ and then with Ar^+ ions. The Ar^+-to-He^+ sputtering yield ratio is 60 [4].

The W specimen not exposed to the beams in the spectrometer was placed in a general-purpose electronic spectrometer (GPES) of original design. This makes it possible to determine the chemical composition of surface by AES. In this device, the specimens can be heated to any desired temperature by either thermal radiation or electron bombardment of its back (nonworking) surface. Under the operating conditions, the pressure was ~10^{-8} Pa. The primary electron energy E_0 in the GPES was varied between 1200 and 1400 eV; the modulation voltage, between 2.5 and 3.0 V. The current beam on the specimen varied from 40 to 80 μA depending on its purity. The electron beam diameter on the specimen was ≈3 mm; the angle of beam incidence, 20°. The GPES was equipped with an originally designed four-grid analyzer of the energy of retarding field electrons [5]. After the specimen had been heated in the interval of 300–2075 K, Auger spectra were recorded in the form of the first derivative of the secondary electron energy distribution.

Under conditions of carbon removal by oxidation, the partial oxygen pressure P_O was 5×10^{-5} Pa; the oxidation time, 2 min at T = 1100 K. The subsequent oxidation to remove metal oxides lasted 1 min at 1900 K (the value of P_O remained the same). The source of spectral-grade atomic oxygen was described in [6]. When the GPES was degassed by heating to 625 K, the specimen remained cool. This allowed us to determine the initial impurity composition and to trace its temperature variation.

SUBSURFACE IMPURITIES

The SIMS spectra show that impurities responsible for the most intense peaks (Na, Al, K, Ca, and Cr) are nonuniformly distributed over the surface (Fig. 1). Other impurities, such as H_2O, C, Si, Ge, and Fe, are distributed more or less uniformly. In addition, the peaks from S, Cl, Ni, Sc, Cu, Zn, Ag, and Pd somewhat exceed the background level. Peaks of many metal oxides are also present. Among them, those from WO and WO_2 are the highest. The peak from WOH is also seen. It is most likely to be associated with water dissociation under the action of the ion beam. A faint peak of WC carbide is also present. The presence of the intense WO peaks against the weak O peak means, first, that W on the surface is severely oxidized and, second, that O is poorly dissolved in CVD tungsten. For the C and WC peaks, the situation is reverse. Ta and Re are impurities attendant to tungsten in the nature.

Initially, the He^+ beam gave rise to rather intense H_2O and F peaks. The former is explained by the specific decontamination conditions prior to placing the specimens into the instrument and also by the fact that they were not heated; the latter, primarily by the CVD conditions. Later, when the Ar^+ beam was used, the F peak rapidly disappeared, and the H_2O peak substantially decreased. Hence, these impurities concentrate largely near the surface. Cr and Fe impurities, conversely, concentrate deep in the coating, since their associated intense peaks appear only during Ar^+ ion bombardment.

The presence of Nb and Zr in some areas on the surface, which are the substrate elements, indicates that either the W coating is discontinuous (although its rated thickness is 30 μm) or Nb and Zr diffuse through it. Note that the presence of Nb on the surface was detected for small exposure times on both specimens in the very beginning of the study; therefore, this fact cannot be attributed to the ion milling of the surface. For the third specimen, which was nondestructively studied with the GPES, the presence of both elements was detected even in the first spectrum, which corresponds to the initial condition of the surface. Its temperature was no more than 625 K when the GPES was degassed.

The exposure of the specimens to the Ar^+ beam for several hours changed the impurity composition and

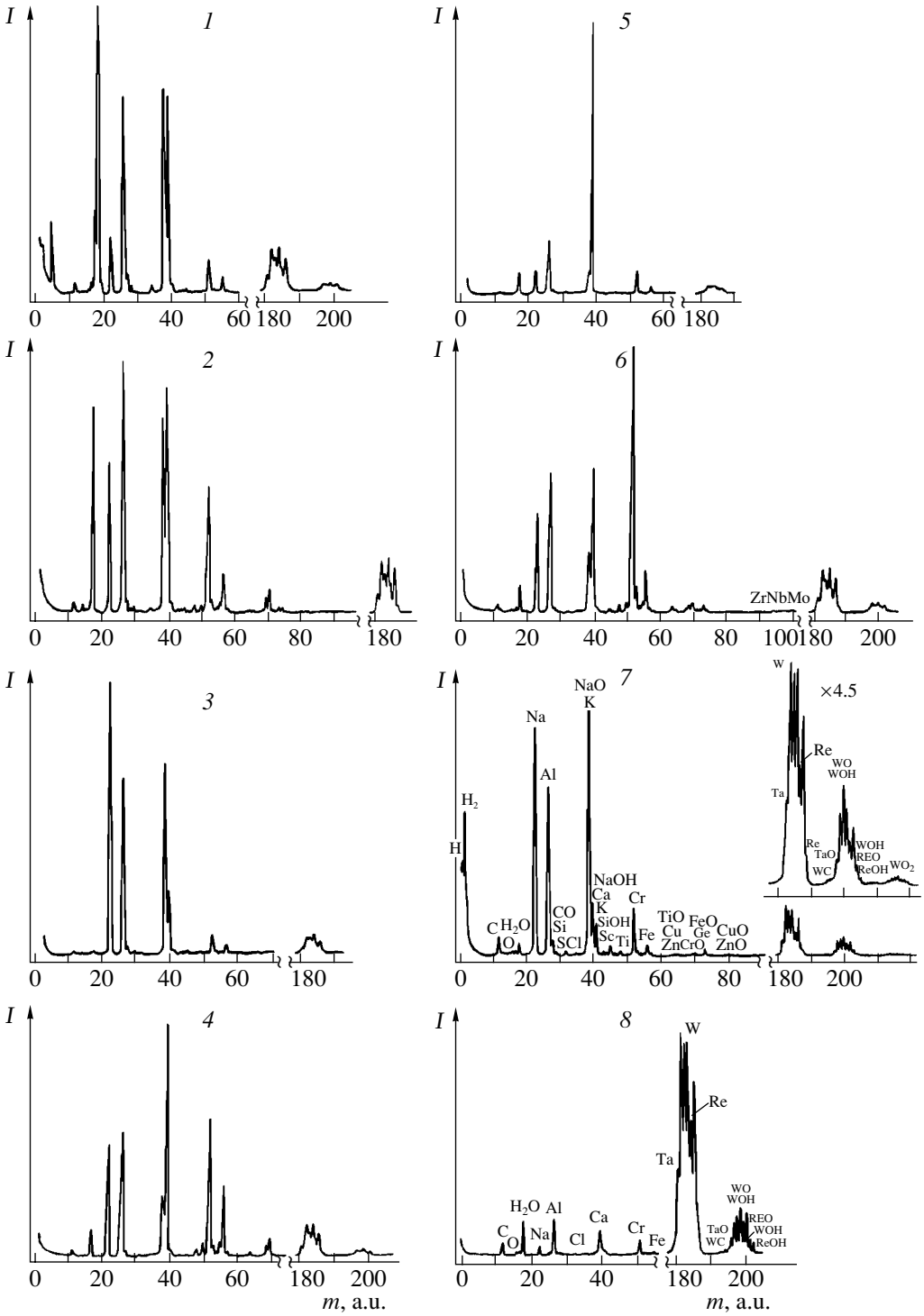


Fig. 1. Mass spectra taken from the W coating on the Nb + 1% Zr substrate. (*1–6*) Refer to areas heavily contaminated by fluorine, water and potassium, sodium, aluminum, calcium, and chromium, respectively, (*7*) the cleanest area, and (*8*) W(110) reference surface.

distribution insignificantly, except for H_2O and F, as was mentioned above. This implies that the W coating is contaminated to a great depth. Among contaminants, there is carbon, which is extremely undesirable.

The mass spectra taken from the W coating were compared with those from the single-crystal W(110) reference under the same conditions. Figure 1 clearly demonstrates that the W(110) surface (curve *8*) is much

less contaminated than the working surface of the TEC collector (curve *7*).

SURFACE IMPURITIES

Auger spectra taken immediately after the instrument has been evacuated exhibit faint peaks of S and W, a more intense peak of O, and a high peak of C (Fig. 2, curve *1*). Then, the surface was decontaminated by heating the specimen from 325 to 2075 K with a step of 50 K. The Auger spectra were recorded both from the cooling-down specimen and at 295 K.

Starting from $T = 425$ K, faint Nb peaks appear at 92 and 120 eV, while the S peak at 152 eV and the W peaks at 168 and 178 eV begin to grow. After heating to 825 K, Ca peaks at 294 and 318 eV appear and at $T = 925$ K, a potassium peak at 253 eV. At $T = 1025$ K, the K peak reaches the maximum intensity, while the C peak at 272 eV totally disappears and the S peak diminishes sharply (Fig. 2, curve *2*). At $T = 1175$ K, the K peak disappears. At $T = 1225$ K, the S peak somewhat declines but additional Nb peaks at 105 and 141 eV arise. At $T = 1275$ K, the S peak disappears (Fig. 2, curve *3*). At $T = 1325$ K, the S peak appears again together with one more Nb peak at 197 eV; however, the major O peak at 512 eV starts lowering. At 1375 K, the Ca peaks are absent, the S peak grows, and the C peak arises again. At 1425 K, the S peak reaches the maximal intensity, exceeding even the intensity of the W peaks, and the Nb peak at 105 eV increases (Fig. 2, curve *4*). The heights of the C and O peaks remain unchanged. At 1525 K, the S peak starts to decrease, while the intensities of all the Nb peaks grow. At $T = 1625$ K, the S peak is absent (Fig. 2, curve *5*). At 1725 K, WC and W_2C carbide peaks at 261 and 272 eV emerge. They continue to grow at 1775 K; however, the Nb and O peaks at this temperature decrease. In the interval of 1875–2075 K, faint Nb and O peaks and intense peaks from tungsten and its carbides are observed (Fig. 2, curve *6*).

Thus, the heating of the W coating triggers the following processes. First, a thick film of amorphous C or hydrocarbons is removed from the surface. Even after heating to 425 K, it becomes clear that the W coating is discontinuous. Subsequently, first Ca and then K diffuse from the interior toward the surface and migrate over it, "screening" deep-seating carbon and sulfur impurities. The C and S impurities are obviously present in the W coating, since Nb is not screened; instead, its peaks grow. Carbon is completely screened by those impurities after heating to 1025 K; sulfur, after heating to 1272 K. It is hard to perceive that Ca and K screen C and S individually; hence, one can argue that S in the surface layer is in greater amount than C. At the same time, of the two impurities diffusing toward the surface, Ca is in greater amount than K, as is also demonstrated by the SIMS data. After heating to 1175 K, potassium is absent on the surface but S is screened as yet incompletely. The screening becomes complete at

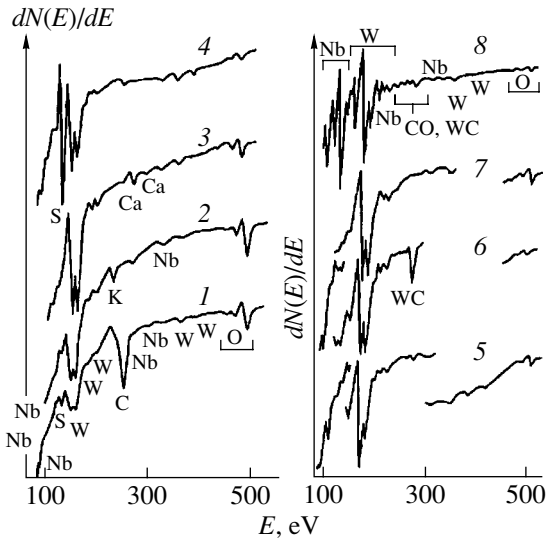


Fig. 2. Auger spectra taken from the W coating on the Nb + 1% Zr substrate. (*1*) As-prepared surface at $T = 295$ K, (*2*) heavy contamination by potassium ($T = 1025$ K), (*3*) screening of sulfur by potassium and calcium ($T = 1275$ K), (*4*) maximal sulfur desorption ($T = 1425$ K), (*5*) surface decontaminated from sulfur ($T = 1625$ K), (*6*) surface heated *in vacuo* at $T = 2075$ K, (*7*) single oxidation at $T = 1125$ K, and (*8*) surface after 50 oxidations and heating at $T = 1675$ K.

1275 K. It is not improbable that S occupies defects in the Nb + 1% Zr substrate, since at 1325 K, when the S peak appears again in the Auger spectrum, the additional Nb peak arises and the O peak diminishes. That is, S diffuses toward the surface and niobium begins to dissolve oxygen. The desorption of oxygen from W, and especially from Nb, is out of the question at such temperatures [7–9]. When Ca desorbs at 1375 K, C and S contaminants become exposed again. Above 1425 K, the S segregation is maximal and sulfur sublimates, completely disappearing at 1625 K. After the S sublimation, the W and Nb carbides start to form, because the carbide peaks at 261 and 272 eV grow, while those of Nb decline. Unfortunately, the amount of oxygen dissolved in the niobium islands turns out to be insufficient to remove carbon from the surface (the O peak lowers). Thus, heating *in vacuo* up to 2075 K fails to remove carbon, the most unwanted impurity for emission electronics, from the surface of the W coating under study. This is because oxygen leaves W at 1875–1975 K [7] and still does not emerge on the surface from Nb–O compounds [8, 9]. Moreover, Nb is present on the W surface in small amounts.

CARBON REMOVAL FROM THE SURFACE

Among the conventional methods for decontaminating the transition metal surface (oxidation, heating in hydrogen, and milling by inert gas ions), oxidation is the most efficient. Initially, oxidation and heating in hydrogen were carried out at relatively high tempera-

tures: 1900–2000 K for heating in H_2 and 2100–2300 K for oxidation in O_2. At these temperatures, metal hydrides and oxides dissociate; that is, the sticking coefficient for these gases equals zero [10]. At the same time, the gas atom lifetime on the surface is other than zero because of an increased gas pressure in the flow and resulting C–H or C–O compounds, i.e., CH_4 or CO, are pumped off by the vacuum system. Such a decontamination procedure requires high temperatures, as was already noted, and takes much time (several hours). If the specimen is heated by electron bombardment, the service life of the heater (electron gun) shortens substantially. Next, since thermal heating lasts long, some parts of the equipment may oxidize. Finally, a number of metals, such as V, W, Ta, Ti, etc., may be heavily sputtered at such temperatures. This adversely affects the insulation of optoelectron and ion–optic systems, electron energy analyzers, monochromators, and other devices integrated into the instrument chamber. Reliable screens around the specimen are frequently impossible to mount, because they are incompatible with an investigation technique.

Therefore, a more appropriate method, cyclic oxidation of carbon or metal carbides, was later suggested [11]. Both the oxygen pressure and the temperature here are maintained at a lower level: $P_O = (1–5) \times 10^{-5}$ Pa and $T \approx 1100$ K. At this temperature, the carbides and metals effectively oxidize, and CO is readily desorbed but oxygen is not desorbed. Therefore, 2 to 3 min after carbide oxidation, the temperature is raised for ≈1 min to remove metal oxides (to $T = 1900$ K for Mo and W). Then, the temperature is reduced to 1100 K again without reducing P_O, and the procedure is repeated. During a cycle, at least one impurity monolayer is removed from the surface. If W and Mo single crystals are perfect, 10 to 15 cycles will suffice to remove carbon from the surface and the decontamination time does not exceed 1 h. After the oxygen pressure has been decreased to the residual pressure, the purity of the surface is checked by AES.

We performed single oxidation following the above technique to estimate the C oxidation efficiency for the polycrystal and then heated the specimen to see whether C is contained inside the specimen. After oxidation at 1125 K, the carbon peak in the Auger spectrum disappears, and Nb and W are screened by oxygen (completely and partially, respectively; Fig. 2, curve *7*). The O peaks grow. However, after heating to 1475 K, the carbide peaks appear again, their intensity being the same as before the oxidation. This means that C diffuses toward the surface and segregates on it at this temperature, while oxygen is not as yet desorbed from W [7]. Fifteen subsequent oxidation–irradiation cycles change the situation observed after the first cycle insignificantly. Therefore, we performed 35 extra oxidation–irradiation cycles. Eventually, after heating to 1525 K, the intensity of the carbide peaks sharply decreased (Fig. 2, curve *8*). In addition, the Nb peaks grow and, after heating to 1625–1675 K, some of them become comparable to the W major peaks. This means that the area occupied by Nb in the coating increased appreciably after such processing. Since the Nb intensity in its Auger spectrum is always three times greater than the Nb intensity in the W spectrum [12], one can suppose that the area occupied by Nb in the W coating may reach 30% after carbon is removed. The fact that heating may increase the surface fraction of a lighter or lower melting alloy component and that this fraction may rise still further after the surface has been exposed to oxygen is well known [13]. In our case, we can state with assurance that the W–Nb interface has the properties of W–Nb alloy.

As the temperature is increased further ($T > 1675$ K), the Nb peaks begin to decrease. It is likely that a part of oxygen leaves Nb for the surface and screens Nb if the metal lies in the W surface recesses. Indeed, after the specimens have been exposed to a vacuum (~10^{-8} Pa) for ≈0.5 h or more (within this time, O and especially CO cannot evolve from the residual gases, since $P_O \sim 10^{-10}$ Pa and the O-to-CO sticking coefficient ratio is ≈10), the Nb peak intensity grows again. Hence, oxygen again dissolves in Nb [8, 9].

With the temperature increased to 2075 K, the C peak intensity differs from that achieved after the decontamination procedure only slightly. Nevertheless, to remove C as completely as possible, we performed 25 additional oxidation–irradiation cycles. They did not cause any changes in the Auger spectrum recorded after 50 preceding cycles. One can conclude, therefore, that the remaining carbon occupies largely deep volume defects (e.g., grain boundaries) and cannot be effectively removed from the W coating within a reasonable decontamination time.

It should also be noted that the W coating is heavily sputtered (which is observed even visually on a shutter placed near the specimen) at temperatures above 1925 K. Such a great amount of the deposit has not been detected even after the completion of the program involving the oxidation of carbon, pure surface, and O–Cs binary films on Mo(110), Nb(110), Mo–Nb(110), and W(110) single crystals and lasting several years. Within this program, a variety of Nb(110)–O, Nb(110)–O–Cs, Mo–Nb(110)–O, and Mo–Nb(110)–O–Cs films were deposited on the atomically clean surface, which was produced by multiply increasing the temperature up to 2500 K. Therefore, severe sputtering of the W coating at $T > 1925$ K may be related to its polycrystallinity (i.e., imperfection) and possibly to the presence of Nb in a certain part of the coating (discontinuity).

CONCLUSION

Thus, almost all unwanted impurities can be removed from a CVD tungsten coating on a Nb + 1% Zr substrate used as the TEC collector material by singly

heating *in vacuo* to $T = 1625$ K. The remaining impurities are Nb, O, and tungsten carbides, the last being the most harmful. The carbides are not removed by heating up to $T = 2075$ K. Moreover, at $T > 1925$ K, the coating is heavily sputtered. However, after 50 oxidation–irradiation cycles, the amount of carbon in the subsurface layers is substantially reduced, although it is not completely removed for a reasonable time. This is because the specimen is strongly contaminated by carbon and also because the W coating is polycrystalline. A great concentration of intrinsic defects favors carbon diffusion toward the surface with subsequent segregation. Therefore, it seems to be natural to make the TEC collector from single-crystalline tungsten in an oxygen-containing atmosphere [14]. To improve the properties of a W coating on Nb + 1% Zr alloy, one has to carefully fulfill the vacuum hygiene requirements at all stages of TEC collector manufacturing. It is imperative here to use hydrocarbon-free pumping facilities. Finally, since at least partial contamination of the collector cannot be avoided, the electrodes rinsed by the technique mentioned above and assembled must be decontaminated immediately before operation. It is best to decontaminate both electrodes simultaneously: first by mere heating and then by the technique described above to remove carbon. The decontamination quality can tentatively be simulated with a special bench provided that all the steps in the electrode fabrication process are exactly reproduced.

As for the other two hard-to-remove impurities, Nb and O, the latter facilitates Cs adsorption [1]. The presence of oxygen on the W surface after heating to temperatures exceeding the temperature of oxygen desorption from tungsten [7] indicates that oxygen is highly likely to occupy surface recesses, entering into the composition of Nb_2W_2O intermetallic suboxide, a chemical analog of Nb_4O suboxide [15]. It is known that Nb_4O suboxide is the optimal Nb oxidation state for Cs in the Nb(110)–O–Cs system [16, 17] and that Nb_2Mo_2O is such a state in the (Mo + 2.5%Nb)(110)–O–Cs system [18, 19]. Yet, the presence of Nb_2W_2O suboxide and its effect on Cs adsorption call for further investigation.

REFERENCES

1. J. Levin and F. Gelhaus, in *Direct Conversion of Heat Energy to Electrical Energy and Fuel Elements* (Akad. Nauk SSSR, Moscow, 1968), Vol. 4, p. 109.
2. B. M. Zykov, A. M. Sabel'nikov, V. K. Tskhakaya, *et al.*, Poverkhnost, No. 4, 65 (1987).
3. B. M. Zykov and Yu. I. Nardaya, Zh. Tekh. Fiz. **65** (4), 150 (1995) [Tech. Phys. **40**, 372 (1995)].
4. *Methods of Surface Analysis*, Ed. by A. W. Czanderna (Elsevier, New York, 1975; Mir, Moscow, 1979).
5. B. M. Zykov, V. K. Tskhakaya, and V. I. Yarygin, USSR Inventor's Certificate No. 1062803, Otkrytiya, Izobret., No. 47, 236 (1983).
6. B. M. Zykov and A. M. Sabel'nikov, Prib. Tekh. Éksp., No. 1, 219 (1991).
7. B. M. Zykov, V. P. Kobyakov, and Yu. I. Nardaya, Vysokochist. Veshchestva, No. 1, 71 (1991).
8. B. M. Zykov, D. S. Ikonnikov, and V. K. Tskhakaya, Fiz. Tverd. Tela (Leningrad) **17**, 3562 (1975) [Sov. Phys. Solid State **17**, 2322 (1975)].
9. B. M. Zykov and A. M. Sabel'nikov, Vysokochist. Veshchestva, No. 5, 123 (1990).
10. J. A. Becker, E. J. Becker, and R. G. Brandes, J. Appl. Phys. **32**, 411 (1961).
11. Y. Viswanath and L. D. Schmidt, J. Chem. Phys. **59**, 4184 (1973).
12. L. E. Davis, N. C. MacDonald, P. W. Palmberg, G. E. Riach, and R. E. Weber, *Handbook of Auger Electron Spectroscopy* (Physical Electronics Industries, Eden Prairie, 1976).
13. J. D. Fast, *Interaction of Metals and Gases* (Academic, New York, 1965; Metallurgiya, Moscow, 1975), Vol. 2.
14. V. P. Kobyakov, Poverkhnost, No. 11, 153 (1991).
15. B. M. Zykov and A. M. Sabel'nikov, Poverkhnost, No. 10, 61 (1988).
16. B. M. Zykov and V. K. Tskhakaya, Zh. Tekh. Fiz. **50**, 1771 (1980) [Sov. Phys. Tech. Phys. **25**, 1035 (1980)].
17. B. M. Zykov, A. M. Sabel'nikov, and V. K. Tskhakaya, Poverkhnost, No. 9, 22 (1990).
18. B. M. Zykov, A. M. Sabel'nikov, and V. K. Tskhakaya, Poverkhnost, No. 2, 21 (1986).
19. B. M. Zykov, Yu. I. Nardaya, and A. M. Sabel'nikov, Vysokochist. Veshchestva, No. 4, 116 (1991).

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1182–1192. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 119–129.

**SURFACES,
ELECTRON AND ION EMISSION**

Calculation of the Electron Escape Function and Medium-Energy Photoelectron Yield from Layer-on-Substrate Composites

L. A. Bakaleĭnikov*, K. Yu. Pogrebitskiĭ*, E. Yu. Flegontova, Yang-Koo Cho**, and Hyun-Min Park****

** Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*
e-mail: bakal@ammp.ioffe.rssi.ru
*** Korea Research Institute of Standards and Science, Taejon, Republic of Korea*
Received January 10, 2002

Abstract—Analytical expressions for the electron escape function and X-ray-induced photoemission yield from layer-on-substrate composites are derived by approximately solving the kinetic equation for medium-energy (several keV) electrons. Relationships obtained by Monte Carlo calculation qualitatively agree with the analytical dependences and make it possible to numerically find parameters entering into them. *© 2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Modern microelectronic technologies, dealing with ultrasmall objects (size-comparable to atomic clusters), necessitate the development of diagnostics techniques allowing for high-resolution determination of object parameters. Today, there are a variety of approaches serving the purpose. In many of them, such as EXAFS and SEXAFS [1], XPS, AES [2], and XIEES [2], a desired signal is produced by electrons. The adequate processing of data obtained with these methods requires correct description of electron transport in the material. However, a quantitative theory of electron emission is as yet in the development stage and electron escape from a material is usually treated in terms of approximate phenomenological formulas.

The probability of an electron escaping a sample depends on the total electron flow through its boundary. The electron flow can be evaluated with a kinetic equation—a linear integro-differential equation that is difficult to solve analytically. Because of this, the differential electron flow density is found by two ways. The first (analytical) approach takes into account electron transport features that enable one to simplify the kinetic equation and find an approximate analytical solution. In the other case, the kinetic equation is solved numerically.

The former approach was used in [4, 5], where analytical expressions for the escape function of low- and medium-energy electrons were found for homogeneous semi-infinite samples.

Results obtained in [4] were verified with numerical simulation. In [6], the escape function was calculated by the Monte Carlo method for a number of homogeneous materials. The applicability domain for the analytical approximations of the escape function was determined by comparing the results obtained in [4] with the calculations. The numerical results were found to be in good agreement with the estimates based on the differential cross sections for elastic and inelastic electron–material interactions.

The numerical approach makes it possible to derive the escape function not only for homogeneous samples but also for those of a more complex composition. Calculations carried out for a thin (several tens of nanometers) layer on a semi-infinite substrate showed that the presence of the layer may affect significantly the depth dependence of the escape function. With the layer present, this dependence is no longer described by the analytical expressions obtained in [4].

The aim of this work is to extend the analytical approach put forward in [2, 5] for layer-on-substrate composites. Our analytical expressions for the escape function and total yield are compared with Monte Carlo calculations.

KINETIC EQUATION AND ESCAPE FUNCTION FOR HOMOGENEOUS SAMPLES

The analytical approach to solving the kinetic equation [4, 5] takes into account the fact that angular and energy relaxations of medium-energy (several keV) electrons have different scales. In the case of a flat-boundary semi-infinite sample, the kinetic equation has the form

$$\cos\Theta\frac{\partial\Phi}{\partial z} = R_{\text{col}}\Phi + \frac{1}{4\pi}\delta(z - z_0)\delta(E - E_0). \qquad (1)$$

1063-7842/02/4709-1182$22.00

Here, $\Phi(z, \Theta, E)$ is the differential flow density of electrons with an energy E moving at a depth z at an angle Θ to the z axis, and $R_{col}\Phi$ is the collision integral, which describes electron–material interaction. The second term on the right means the presence of a source of electrons with an energy E_0 at a depth z_0. Equation (1) should be supplemented by the boundary conditions

$$\Phi(0, \Theta, E)|_{\cos\Theta > 0} = 0, \quad \Phi(z, \Theta, E)|_{z \to \infty} \longrightarrow 0,$$

which mean the absence of an electron flow from the outside and the absence of electrons at large depths. The probability that an electron with an energy E_0 escapes the sample from a depth z_0 (escape function) depends on the total electron flow through the surface $z = 0$:

$$q(E_0, z_0) = \int_0^{E_0} \int_0^{2\pi} \int_0^{\pi} \cos\Theta \Phi(0, \Theta, E) \sin\Theta d\Theta d\varphi dE. \quad (2)$$

The characteristic lengths of electron transport in a material are the mean free paths for elastic,

$$l_{el}(E) = (N\sigma_{el})^{-1} = \left(N 2\pi \int_0^{\pi} \frac{d\sigma_{el}(E, \Theta)}{d\Theta} \sin\Theta d\Theta \right)^{-1}$$

and inelastic,

$$l_{in}(E) = (l_{in}^{-1}(E))^{-1} = \left(\int_0^{Q_{max}} \frac{dl_{in}^{-1}}{d\hbar\omega} d\hbar\omega \right)^{-1}$$

collisions, where

$$\frac{d\sigma_{el}(E, \Theta)}{d\Theta}$$

is the elastic scattering differential cross section, N is the density of scatterers, $dl_{in}^{-1}/d\hbar\omega$ is the inverse differential mean free path (the probability that an electron with an energy E loses an energy $\hbar\omega$ within a unit length), and Q_{max} is the maximum possible energy loss.

Since the energy loss and a change in the angle in an elementary electron–medium interaction event are small, the relaxation process as a whole is characterized by the transport mean free path and the total range of an electron in the medium rather than by the free paths $l_{el}(E)$ and $l_{in}(E)$. As scales for the transport path and electron range, one can take the quantities $\lambda(E)$ and $S(E)$ defined as

$$\lambda(E) = 2\left(2\pi N \int_0^{\pi} \frac{d\sigma_{el}(E, \Theta)}{d\Theta} \Theta^2 \sin\Theta d\Theta \right)^{-1},$$

$$S(E) = E/\beta(E),$$

where $\beta(E)$ is the mean energy loss per unit length:

$$\beta(E) = \int_0^E \hbar\omega (dl_{in}^{-1}/d\hbar\omega) d\hbar\omega.$$

The angular redistribution upon electron–material interaction is specified largely by elastic scattering; hence, the collision integral $R_{col}\Phi$ can be represented as the sum of the elastic ($R_{el}\Phi$) and inelastic ($R_{in}\Phi$) components:

$$R_{el}\Phi = N \int_{4\pi} \frac{d\sigma_{el}(E, \mathbf{\Omega'\Omega})}{d\Omega} \Phi(z, \mathbf{\Omega'}, E) d\mathbf{\Omega'}$$

$$- N\Phi(z, \mathbf{\Omega}, E) \int_{4\pi} \frac{d\sigma_{el}(E, \mathbf{\Omega\Omega'})}{d\mathbf{\Omega}} d\mathbf{\Omega'},$$

$$R_{in}\Phi = \int_0^{E_0} \frac{dl_{in}^{-1}(E + \hbar\omega, \hbar\omega)}{d(\hbar\omega)} \Phi(z, \mathbf{\Omega}, E + \hbar\omega)$$

$$- \Phi(z, \mathbf{\Omega}, E) l_{in}^{-1}(E).$$

Here, $\mathbf{\Omega}$ and $\mathbf{\Omega'}$ are the directions of electron motion before and after scattering. It was shown [7] that for medium-energy electrons propagating in heavy targets (with the atomic number $Z > 50$), the relationship

$$\lambda(E) \ll S(E) \quad (3)$$

is valid. This means that the processes of angular and energy relaxations can be separated and the motion of electrons in a material (on a large scale) is the diffusion of isotropically distributed electrons that is accompanied by energy loss. As was demonstrated in [5], inequality (3) allows one to separate out the smallness parameter $\varepsilon = \sqrt{\lambda(E_0)/S(E_0)}$ from kinetic equation (1) (with $\sqrt{\lambda(E_0)S(E_0)}$ and E_0 as the length and energy scales) and take advantage of methods from the perturbation theory. In this case, the kinetic equation is simplified to an equation for the isotropic component, which depends only on depth and energy. The resulting equation can be simplified further if

$$l_{in}(E) \ll S(E).$$

In this case, one can employ the continuous slowing-down approximation, in which the inelastic part of the collision integral is approximated by the differential operator

$$R_{in}\Phi = \partial(\beta(E)\Phi)/\partial E.$$

As a result, the kinetic equation for the differential flow density is reduced to the diffusion equation. In dimensionless variables, this equation for a homogeneous sample is given by [5]

$$\frac{1}{3\nu_1}\frac{\partial^2 U_0}{\partial z^2} - \frac{\partial}{\partial E}(\gamma(E)U_0) = \delta(z - z_0)\delta(E - 1). \quad (4)$$

Here, $U_0(z, E)$ is the isotropic part of the distribution, which depends only on depth and energy, $\gamma(E) = S(E_0)\beta(E)$, and the energy E is measured in units of E_0. The differential flow density $\Phi(z, \Theta, E)$ is related to $U_0(z, E)$ as

$$\Phi(z, \Theta, E) = U_0(z, E) + \sqrt{\frac{\lambda(E_0)}{S(E_0)}}\frac{\cos\Theta}{\nu_1(E)}\frac{\partial U_0(z, E)}{\partial z} + O(\varepsilon^2), \quad (5)$$

where $\nu_1(E)$ is the first eigenvalue of the dimensionless elastic part of the collision integral.

As was noted in [5], the eigenfunctions of the operator $R_{el}\Phi$ are spherical harmonics $Y_{i,k}(\Omega)$; hence,

$$\lambda R_{el} Y_{1,0} = \nu_1 Y_{1,0}.$$

The solution to Eq. (4) has the form

$$U_0(z, E) = \frac{1}{\gamma(E)}\frac{1}{8\pi^{3/2}\tau(E)}(\exp(-(z - z_0)^2/(4\tau(E))) - \exp(-(z + z_0)^2/(4\tau(E)))),$$

where

$$\tau(E) = \int_1^E \frac{dE}{3\gamma(E)\nu_1(E)}.$$

Substituting this solution into (5) and (2) yields an analytical expression for the escape function. It should be noted that the diffusion approximation is valid only away from the surface (at depths comparable to or greater than the isotropization length). Therefore, for the problem to be solved correctly, it is necessary to consider the boundary layer approximation. The boundary condition for the diffusion approximation is found by sewing together both solutions. This leads to the appearance of some energy-dependent factor C in the expression for the escape function, namely,

$$q(z_0) = C\left[1 - \mathrm{erf}\left(\frac{z_0}{L}\right)\right]. \quad (6)$$

In practice, factor C can be found by numerical calculation.

The escape length L entering into (6) turns out to be related to the lengths of angular and energy relaxations as

$$L(E_0) = 2\sqrt{\lambda(E_0)S(E_0)\tau^*}, \quad (7)$$

where $\tau^* = \tau(0)$ is a dimensionless coefficient.

ESCAPE FUNCTION FOR LAYER-ON-SUBSTRATE COMPOSITES

For layer-on-substrate composites, the electron kinetics is characterized by two lengths describing electron–layer and electron–substrate interactions. Let the characteristic transport mean free paths of electrons in the layer and in the substrate (isotropization length, energy relaxation length, and the first eigenvalue of the elastic part of the collision integral) be denoted as $\lambda^{(1)}(E)$, $S^{(1)}(E)$, $\nu_1^{(1)}(E)$ and $\lambda^{(2)}(E)$, $S^{(2)}(E)$, $\nu_1^{(2)}(E)$, respectively.

In going to dimensionless variables in Eq. (1), the quantity $\sqrt{\lambda^{(1)}(E_0)S^{(1)}(E_0)}$ for the layer is taken as the length scale. Consider the case when an electron source is in the layer at a depth $z_0 < a$, where a is the thickness of the layer. In terms of the notation introduced above, the differential equation for the isotropic part of the distribution takes the form

$$\frac{1}{3\nu_1^{(1)}(E)}\frac{\partial^2 U_0}{\partial z^2} - \frac{\partial}{\partial E}(\gamma^{(1)}(E)U_0) = \delta(z - z_0)\delta(E - 1),$$

for $0 < z < a$,

$$\frac{\lambda^{(2)}(E_0)}{\lambda^{(1)}(E_0)}\frac{1}{3\nu_1^{(2)}(E)}\frac{\partial^2 U_0}{\partial z^2} - \frac{S^{(1)}(E_0)}{S^{(2)}(E_0)}\frac{\partial}{\partial E}(\gamma^{(1)}(E)U_0) = 0,$$

for $a < z < \infty$.

Designating $B_0 = \gamma^{(1)}U_0$, we find

$$\frac{1}{3\nu^{(1)}(E)\gamma^{(1)}(E)}\frac{\partial^2 B_0}{\partial z^2} - \frac{\partial B_0}{\partial E} = \delta(z - z_0)\delta(E - 1),$$

for $0 < z < a$,

$$\frac{\lambda^{(2)}(E_0)S^{(2)}(E_0)}{\lambda^{(1)}(E_0)S^{(1)}(E_0)}\frac{1}{3\nu^{(2)}(E)\gamma^{(1)}(E)}\frac{\partial^2 B_0}{\partial z^2} - \frac{\partial}{\partial E}\left(\frac{\gamma^{(2)}(E)}{\gamma^{(1)}(E)}B_0\right) = 0, \quad (8)$$

for $a < z < \infty$.

This equation can be simplified further if it is taken into account that $\gamma^{(2)}(E)/\gamma^{(1)}(E)$ is close to unity and weakly depends on the energy. In fact, if the mean energy loss is given by the Bethe formula, this ratio has

the form

$$\frac{\gamma^{(2)}(E)}{\gamma^{(1)}(E)} = \left(1 + \frac{\sum_i N_i^{(2)} Z^{(2)} \ln(E)}{\sum_i N_i^{(2)} Z_i^{(2)} \ln(1.166E_0/I_i^{(2)})}\right) \Bigg/ \left(1 + \frac{\sum_i N_i^{(1)} Z^{(1)} \ln(E)}{\sum_i N_i^{(1)} Z_i^{(1)} \ln(1.166E_0/I_i^{(1)})}\right).$$

Here, $N_i^{(1)}$ and $N_i^{(2)}$ are the densities of type-i scatterers in the layer and the substrate, respectively; $Z_i^{(1)}$ and $Z_i^{(2)}$ are the atomic numbers of the scatterers; and $I_i^{(1)}$ and $I_i^{(2)}$ are the mean ionization potentials. In the first parentheses, summation is over all substrate components; in the second, over all components of the layer.

It is seen that $\gamma^{(2)}(E)/\gamma^{(1)}(E) = 1 + O(1/\ln(E_0/I))$, where I is the effective ionization potential of the target. In the zero approximation, this ratio can be set equal to unity. Then, for the case $a < z < \infty$, Eq. (8) becomes

$$\frac{\lambda^{(2)}(E_0)S^{(2)}(E_0)}{\lambda^{(1)}(E_0)S^{(1)}(E_0)} \frac{1}{3\nu_1^{(2)}(E)\gamma^{(2)}(E)} \frac{\partial^2 B_0}{\partial z^2} - \frac{\partial B_0}{\partial E} = 0.$$

Changing variables,

$$\tau = \int_1^E \frac{dE}{3\nu_1^{(1)}(E)\gamma^{(1)}(E)},$$

one easily obtains

$$\frac{\partial^2 B_0}{\partial z^2} - \frac{\partial B_0}{\partial \tau} = \delta(z - z_0)\delta(E - 1), \quad z < a,$$
$$D\frac{\partial^2 B_0}{\partial z^2} - \frac{\partial B_0}{\partial \tau} = 0, \quad z > a, \tag{9}$$

where

$$D = \frac{\lambda^{(2)}(E_0)S^{(2)}(E_0)}{\lambda^{(1)}(E_0)S^{(1)}(E_0)} \frac{\nu_1^{(1)}(E)\gamma^{(1)}(E)}{\nu_1^{(2)}(E)\gamma^{(2)}(E)}.$$

Now consider the ratio $\nu_1^{(2)}(E)/\nu_1^{(1)}(E)$. With the Rutherford screened cross section used to describe elastic collisions, this ratio has the form

$$\frac{\nu^{(2)}(E)}{\nu^{(1)}(E)} = \left(1 - \frac{\sum_i N_i^{(2)} (Z^{(2)})^2 \ln(E)}{\sum_i N_i^{(2)} (Z_i^{(2)})^2 \ln(\alpha_i^{(2)}(E_0))}\right) \Bigg/ \left(1 - \frac{\sum_i N_i^{(1)} Z^{(1)} \ln(E)}{\sum_i N_i^{(1)} Z_i^{(1)} \ln(\alpha_i^{(1)}(E_0))}\right),$$

where $\alpha_i^{(k)}(E_0) = (0.565(Z_i^{(k)})^{1/3} e^2 \sqrt{m}/\hbar\sqrt{2E_0})^2$ ($k = 1$, 2) and m and e are the mass and charge of an electron.

In the first parentheses, summation is over all substrate components; in the second, over all components of the layer. Taking into account that $\alpha_i^{(k)}(E_0) \ll 1$ for any i and k, one can conclude that $\nu_1^{(2)}(E)/\nu_1^{(1)}(E) = 1 + O(1/\ln\alpha)$, where α is the maximal value of $\alpha_i^{(k)}$. Therefore, we will assume hereafter that $\nu_1^{(2)}(E)/\nu_1^{(1)}(E) = 1$. For D, we have

$$D = \left(\frac{L^{(2)}}{L^{(1)}}\right), \tag{10}$$

where $L^{(1)}$ and $L^{(2)}$ are the escape lengths in the layer and in the substrate, respectively, which are defined by relationship (7).

To find a solution to (9), this equation should be supplemented by boundary conditions. The absence of electron sources at $z = 0$ and $z \longrightarrow \infty$ is written as

$$U_0|_{z=0} = 0, \quad U_0|_{z\to\infty} = 0.$$

Since electrons can only lose the energy and sources of electrons with an energy higher than E_0 are absent,

$$U_0|_{E=1+0} = 0.$$

Passing from U_0 to the function B_0 and using the age of the particle τ instead of the dimensionless energy E,

we get

$$B_0(0,\tau) = 0, \quad B_0(z,\tau)|_{z\to\infty} = 0, \quad B_0(z,0) = 0. \tag{11}$$

The problem defined by (9) and (11) can be solved with the Laplace transformation. The transform $F(z, p)$ satisfies the equation

$$\frac{\partial^2 F(z,p)}{\partial z^2} - pF(z,p) = \delta(z-z_0), \quad z<a,$$
$$D\frac{\partial^2 F(z,p)}{\partial z^2} - pF(z,p) = 0, \quad z>a. \tag{12}$$

To calculate the escape function, it is necessary to find a solution to (12) in the interval $z < z_0$. One can show that this solution has the form

$$F(z,p) = -\frac{1}{2\sqrt{p}} \times \frac{\exp(-\sqrt{p}z_0) + \frac{(\sqrt{D}-1)}{(\sqrt{D}+1}\exp(-\sqrt{p}(-z_0+2a))}{1-\left(\frac{(\sqrt{D}-1)}{(\sqrt{D}+1}\right)\exp(-\sqrt{p}2a)} \times (\exp(\sqrt{p}z) - \exp(-\sqrt{p}z)). \tag{13}$$

Before calculating the original, we note that τ in (9) is bounded above: $\tau < \tau^*$. If the Bethe formula and the Rutherford screened cross section are used, τ^* roughly equals 1/12. Thus, we can consider only small τ (or large p). Then, Eq. (13) can be expanded in the small parameter $\exp(-\sqrt{p}\,2a) \ll 1$, which yields

$$F(z,p) = -\frac{1}{2\sqrt{p}}\{\exp(-\sqrt{p}(-z+z_0)) + \exp(-\sqrt{p}(z+z_0)) + \sum_{k=1}^{\infty}\left(\frac{\sqrt{D}-1}{\sqrt{D}+1}\right)^k \times [\exp(-\sqrt{p}(-z+z_0+2ka)) - \exp(-\sqrt{p}(z+z_0+2ka)) - \exp(-\sqrt{p}(-z+z_0+2ka)) + \exp(-\sqrt{p}(z+z_0+2ka))]\}.$$

Passing to the original, we find the solution to (9) in the interval $z < z_0$:

$$B_0(z,\tau) = \frac{1}{2\pi\sqrt{\tau}}\left\{\exp\left(-\frac{(-z+z_0)^2}{4\tau}\right) - \exp\left(-\frac{(z+z_0)^2}{4\tau}\right) + \sum_{k=1}^{\infty}\left(\frac{\sqrt{D}-1}{\sqrt{D}+1}\right)^k \times\left(\exp\left(-\frac{(-z+z_0+2ka)^2}{4\tau}\right) - \exp\left(-\frac{(z+z_0+2ka)^2}{4\tau}\right) - \exp\left(-\frac{(-z-z_0+2ka)^2}{4\tau}\right) + \exp\left(-\frac{(z-z_0+2ka)^2}{4\tau}\right)\right)\right\}. \tag{14}$$

The first two terms in (14) give the solution for a homogeneous sample [4]. The effect of the substrate on the differential flow density in the layer is reflected by summation over $k \neq 0$.

Similar considerations for the case where electron sources are in the substrate result in

$$B_0(z,\tau) = \frac{1}{\sqrt{\pi}(1+\sqrt{D})}\frac{1}{\sqrt{\tau}}\sum_{k=0}^{\infty}\left(\frac{\sqrt{D}-1}{\sqrt{D}+1}\right)^k \times\left(\exp\left(-\frac{1}{4\tau}\left(\frac{z_0-a}{\sqrt{D}}+(2k+1)a-z\right)^2\right) - \exp\left(-\frac{1}{4\tau}\left(\frac{z_0-a}{\sqrt{D}}+(2k+1)a+z\right)^2\right)\right). \tag{15}$$

To find the escape function for the layer and the substrate, expressions (14) and (15) should be substituted into (5) and (2). The escape function for the layer (accurate to the factor C, as for a homogeneous sample) has the form

$$q(z,E) = C\left\{1-\mathrm{erf}\left(\frac{z}{L}\right) + \sum_{k=1}^{\infty}\left(\frac{\sqrt{D}-1}{\sqrt{D}+1}\right)^k \times\left(\mathrm{erf}\left(\frac{2ka-z}{L}\right) - \mathrm{erf}\left(\frac{2ka+z}{L}\right)\right)\right\}. \tag{16}$$

The first two terms in (16) that do not appear under the summation sign give the escape function for a homogeneous sample. The escape function for the sub-

strate is given by

$$q(z, E) = C\left\{\frac{2}{1+\sqrt{D}}\sum_{k=0}^{\infty}\left(\frac{\sqrt{D}-1}{\sqrt{D}+1}\right)^k \times\left(1-\operatorname{erf}\left(\frac{z-a}{\sqrt{D}L}+\frac{(2k+1)a}{L}\right)\right)\right\}. \quad (17)$$

It is easy to check that (16) and (17) become the escape function for a homogeneous sample in the limit $D \longrightarrow 1$.

NUMERICAL CALCULATION OF THE ESCAPE FUNCTION FOR LAYER-ON-SUBSTRATE COMPOSITES

Expressions (16) and (17) have been derived under two assumptions. The first one is the validity of inequality (3), that is, of the diffusion approximation. The second one is that the ratios $\gamma^{(2)}(E)/\gamma^{(1)}(E)$ and $\nu_1^{(2)}(E)/\nu_1^{(1)}(E)$ (hence, the coefficient D) are energy-independent. These assumptions are valid in a certain range of energies and atomic numbers. The limits of this range can roughly be estimated from the scattering differential cross sections and characteristic lengths.

Another way of finding the applicability domain of analytical expressions for the escape function is to compare them with results of Monte Carlo numerical simulation. In a number of cases, analytical expressions give proper results even if the assumptions made in the analysis fail. In this case, the true values of parameters entering into these expressions should be found by using numerical calculation.

In [7], we described the Monte Carlo method used to characterize the physical model of electron–material interaction, namely, the single scattering model, where each elementary interaction event depends on the elastic and inelastic scattering differential cross sections. In the case of elastic scattering, the Mott differential cross section is employed. Inelastic electron–material interaction is described in terms of the doubly differential inverse free path $d^2 l_{\text{in}}^{-1}/d(\hbar\omega)d(\hbar q)$, which is the probability that an electron with an energy E loses an energy $\hbar\omega$ and a momentum $\hbar q$ within a unit length. The doubly differential inverse free path can be found from data for the material permittivity [8–10]. The calculation of this quantity, as well as the draw procedure as applied to inelastic interaction results that allows for a substantial saving of computational resources, is described in [1]. Differential cross sections for elastic and inelastic interactions are available from the electron archive [12].

In this work, we calculated the escape functions for samples with various layer-to-substrate atomic number ratios (atomic number "contrast"): an aluminum layer on a gold substrate, Al/Au; gold on aluminum, Au/Al; aluminum on copper, Al/Cu; and copper on aluminum, Cu/Al. The thickness of the layer was varied from several nanometers to that comparable to the characteristic depth of electron escape (on the order of several hundreds of nanometers).

The data for the elastic scattering differential cross sections suggest that the layer thickness of about several tens of nanometers is comparable to the characteristic length of isotropization; that is, electrons moving in the layer have no time to be redistributed over angles. Therefore, the escape lengths for the layer, L, appearing in (16) and (17) differ from those for homogeneous samples of the same composition, $L^{(1)}$, and should be considered as depending on the layer thickness.

Corrected escape lengths for the layer L and the factors C were determined by minimizing the functional

$$I(C, L; E, a) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[q_{\text{MC}}(z_i, E, a) - q_{\text{theor}}(z_i, E, a)]^2}, \quad (18)$$

where $q_{\text{theor}}(z_i, E, a)$ is defined by (16) or (17) and z_i are points where the escape function $q_{\text{MC}}(z, E, a)$ is found by the Monte Carlo method.

The values of L thus obtained were substituted into (16) and (17), where they appear both in explicit form and through the coefficient D. In calculating D by formula (10), the escape length for the layer $L^{(1)}$ was replaced by L. For the substrate, the diffusion approximation is valid; therefore, formula (10) uses the escape length for homogeneous samples as the escape length for the substrate, $L^{(2)}$.

Analytical expressions (16) and (17), including the corrected escape lengths L, are in good agreement with the Monte Carlo results for energies of 0.5–10 keV and layer thicknesses of 5–10 nm. For these parameter ranges, typical values of functional (18) are on the order of several thousandths.

Figure 1 shows the escape functions found analytically and numerically for the Al/Au samples at an initial electron energy of 10 keV. The analytical and numerical curves are seen to be nearly coincident.

It is important to estimate the dependence of corrections to the escape length for the layer on electron energy, layer thickness, and layer-to-substrate atomic number ratio and compare it with the same dependence of corrections to the escape length for a homogeneous sample. It was found that the escape length for a layer of a low-Z material on a high-Z substrate decreases compared with that for the homogeneous sample, while in the case of a high-Z layer on a low-Z substrate, the escape length increases. This tendency is aggravated with decreasing layer thickness and growing atomic number contrast. Figure 2 demonstrates the ratios of the escape lengths L for a 10- and 60-nm-thick Al layer in

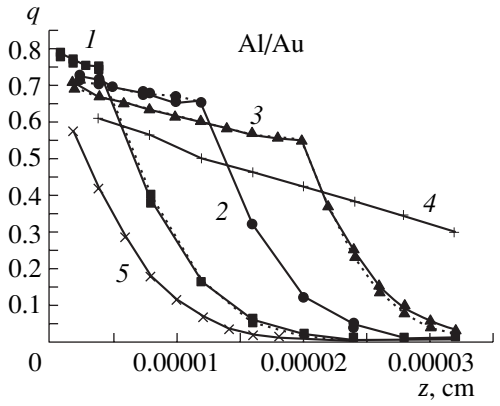


Fig. 1. Depth dependence of the escape function for the Al/Au composite. Solid curves, numerical calculation; dashed curves, analysis. a = 0.000004 (*1*), 0.000012 (*2*), and 0.000020 (*3*). (*4*) Bulk Al and (*5*) bulk Au.

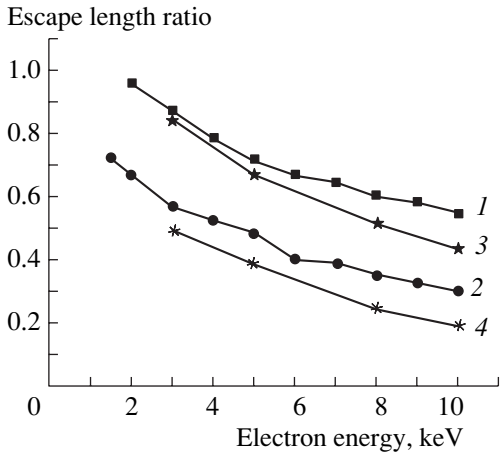


Fig. 2. Ratios of the escape lengths for the layer to those for the semi-infinite Al sample in the case of the Al/Cu and Al/Au composites with a layer thickness of 10 and 60 nm. (*1*) Al/Cu, a = 600; (*2*) Al/Cu, a = 100; (*3*) Al/Au, a = 600; and (*4*) Al/Au, a = 100.

the Al/Au and Al/Cu structures to the escape length $L^{(1)}$ for the homogeneous sample vs. electron initial energy.

CALCULATION OF PHOTOEMISSION FROM LAYER-ON-SUBSTRATE COMPOSITES

The results obtained make it possible to find analytical formulas for calculating electron photoemission from layer-on-substrate composites. Electron emission can be described within the three-stage model. According to it, X-ray-induced electron emission proceeds in three stages: the absorption of an X-ray photon by material atoms (attenuation of X-ray radiation), the generation of primary electrons (i.e., photoelectrons and electrons due to the relaxation of photoionized atoms; atomic photoelectric effect), and the propagation of the primary electrons in the material with the subsequent generation of a secondary-electron cascade and the escape of the resulting electrons from the sample. It is assumed that the three stages proceed independently; that is, the probability of an electron appearing in free space after the incidence of an X-ray photon on the surface is determined as the product of the related probabilities.

The first two stages can be simulated in terms of the radiation transfer equation and energy-level transition probabilities. The latter have been calculated by many teams of researchers in various quantum-mechanical approximations and can be considered known [13]. The X-ray intensity at a depth z is given by

$$I(z) = I_0 \exp\left(-\int_0^z \mu(h\nu, \tau)d\tau/\sin\varphi\right).$$

Here, I_0 is the incident intensity, $\mu(h\nu, z)$ is the linear absorption coefficient, and φ is the angle between the propagation direction of the beam and the surface it strikes.

The third stage (electron transport in the material) is characterized by the escape function $q(E, z)$. The total yield of the primary electrons is expressed as

$$\kappa_p = \int_0^\infty I(z)\sum_i\sum_j \mu_i(z)P_{ij}(z)q(E_{ij}, z)dz/\sin\varphi. \quad (19)$$

Here, μ_i is the linear coefficient of absorption by atoms of the ith element, and P_{ij} is the probability of generating an electron with an energy E_{ij} when an atom of the ith element absorbs an X-ray photon. The contributions of slow and fast electrons generated in the layer and in the substrate to the total yield should be considered separately, because the escape functions differ in these four cases. Note, however, that when the layer thickness is about ten nanometers or more, slow electrons generated in the substrate make a negligible contribution to the total yield.

It was shown [4] that the electron escape function for homogeneous samples is well approximated by formula (6) at energies in the kiloelectronvolt range. For energies of about several tens of electronvolts, the escape function obeys the exponential law

$$q(E_{ij}, z) = C_{ij}\exp\left(-\frac{z}{L_{ij}}\right). \quad (20)$$

At intermediate energies, expressions (20) and (6) approximate the escape function with roughly the same accuracy. Since the effect of the substrate on the escape function is negligible for electron energies of several tens to several hundreds of electronvolts and a layer thickness of 10 nm or more, formula (20) can be taken as the escape function for slow electrons. Substituting

(20) into (19), one readily comes to

$$\kappa_p^{(\mathrm{slow})}(h\nu) = I_0 \sum \frac{L_{ij}^{(1)}\mu_i^{(1)}(h\nu)P_{ij}^{(1)}(h\nu)C_{ij}}{L_{ij}^{(1)}\mu^{(1)}(h\nu) + \sin\varphi} \times \left(1 - \exp\left(-\frac{\mu^{(1)}(h\nu)a}{\sin\varphi} - \frac{a}{L_{ij}^{(1)}}\right)\right) + I_0 \sum \frac{L_{ij}^{(s)}\mu_i^{(s)}(h\nu)P_{ij}^{(s)}(h\nu)C_{ij}}{L_{ij}^{(s)}\mu^{(s)}(h\nu) + \sin\varphi} \times \exp\left(-\frac{\mu^{(1)}(h\nu)a}{\sin\varphi} - \frac{a}{L_{ij}^{(1)}}\right). \quad (21)$$

Here, the quantities with the superscripts (1) and (*s*) refer, respectively, to the layer and the substrate and summation is over all groups of slow electrons. The contribution of fast electrons to the total yield is easily found by substituting (16) and (17) into (19). The contribution from the layer is given by

$$\kappa_p^{(l)} = \int_0^a I(z)\sum_i\sum_j \mu_i^{(l)}P_{ij}^{(l)}q^{(l)}(E_{ij}, z)dz/\sin\varphi$$
$$= I_0 \sum P_{ij}^{(l)}\frac{\mu_i^{(l)}}{\mu^{(l)}}C_{ij}\left\{1 - \exp(-ca)\left(1 - \mathrm{erf}\left(\frac{a}{L_{ij}}\right)\right)\right.$$
$$- \exp(\chi_{ij}^2)\left(\mathrm{erf}\left(\frac{a}{L_{ij}} + \chi_{ij}\right) - \mathrm{erf}(\chi_{ij})\right)$$
$$+ \sum_{k=1}^{\infty}\left(\frac{\sqrt{D_{ij}}-1}{\sqrt{D_{ij}}+1}\right)^k\left[\exp(-ca)\left(\mathrm{erf}\left(\frac{(2k+1)a}{L_{ij}}\right)\right.\right.$$
$$\left.\left. - \mathrm{erf}\left(\frac{(2k-1)a}{L_{ij}}\right)\right) - \exp(\chi_{ij}^2 + 2kac)\right.$$
$$\times\left(\mathrm{erf}\left(\frac{(2k+1)a}{L_{ij}} + \chi_{ij}\right) - \mathrm{erf}\left(\frac{2ka}{L_{ij}} + \chi_{ij}\right)\right)$$
$$+ \exp(\chi_{ij}^2 - 2kac)\left(\mathrm{erf}\left(\frac{(2k+1)a}{L_{ij}} - \chi_{ij}\right)\right.$$
$$\left.\left.\left. - \mathrm{erf}\left(\frac{2ka}{L_{ij}} - \chi_{ij}\right)\right)\right]\right\}. \quad (22)$$

Here, $\mu^{(1)}$, $\mu_i^{(1)}$, and $P_{ik}^{(1)}$ refer to the layer; $c = \mu^{(1)}/\sin\varphi$; and $\chi_{ij} = CL_{ik}/2$. The contribution of primary electrons generated in the substrate to the total yield of photoemission is expressed as

$$\kappa_p^{(s)} = I_0\exp(-\mu^{(1)}a)\sum P_{ij}^{(s)}\frac{\mu_i^{(s)}}{\mu^{(s)}}C_{ij}\frac{2}{1+\sqrt{D_{ij}}}$$
$$\times\sum_{k=0}^{\infty}\left(\frac{\sqrt{D_{ij}}-1}{\sqrt{D_{ij}}+1}\right)^k\left\{1 - \mathrm{erf}\left(\frac{(2k+1)a}{L_{ik}}\right)\right.$$
$$\left. - \exp(D_{ij}\chi_{ij}^2 + (2k+1)ac\sqrt{D_{ij}})\right.$$
$$\left.\times\left(1 - \mathrm{erf}\left(\frac{(2k+1)a}{L_{ik}} + \sqrt{D_{ij}}\chi_{ij}\right)\right)\right\}. \quad (23)$$

Here, $\mu^{(1)}$, $\mu_i^{(1)}$, and $P_{ik}^{(1)}$ refer to the substrate and $c = \mu^{(s)}/\sin\varphi$. Summation in (22) and (23) is over all groups of fast electrons generated in the layer and in the substrate, respectively.

The total contribution of the primary electrons is the sum of all partial contributions:

$$\kappa_p = \kappa_p^{(l)} + \kappa_p^{(s)} + \kappa_p^{(\mathrm{slow})}.$$

The total emission yield (TEY) was calculated by formulas (21)–(23) for a number of homogeneous samples and layer-on-substrate composites. The curves calculated were compared with those obtained by the Monte Carlo numerical simulation of photoemission.

The comparison between the analytical and numerical results on the TEY was made for homogeneous samples of aluminum (the angle of irradiation measured from the surface was $\varphi = 5°$) and germanium ($\varphi = 2°$). The photon energy was varied near the *K* edge of aluminum from 1.5 to 2.0 keV and near the *K* edge of germanium from 11 to 12 keV. The parameters used for the Monte Carlo simulation (linear coefficients of X-ray absorption, photoionization cross sections for individual atomic shells, and probabilities of energy-level transitions) were the same as in analytical expressions for the TEY.

The statistical error of the numerical simulation was about 1%. To attain such an accuracy, we simulated as many as 10^6 events of photon incidence for each of the photon energies. For the aluminum (Fig. 3a), the analytical dependence and the Monte Carlo simulation (only the primary electrons were taken into account) diverge by about 2% at the *K* edge. At a distance of 0.2 eV or more from the *K* edge, both curves coincide within the accuracy of the numerical calculation. For the germanium (Fig. 3b), the discrepancy is about 3%. Partially, it is explained by the fact that the Monte Carlo simulation includes the contribution of fluorescent photons to the TEY. Without fluorescent photons, the deviation between the analytical and numerical results is about 1%, i.e., exceeds the numerical calculation accuracy only slightly. The reason for such a discrepancy is likely to be the insufficiently high accuracy of the analytical approximation. In the case of the aluminum, fluorescent photons do not influence the results.

The contribution of the secondary electrons to the TEY merits special consideration. In Monte Carlo simulation, their contribution may be included as follows:

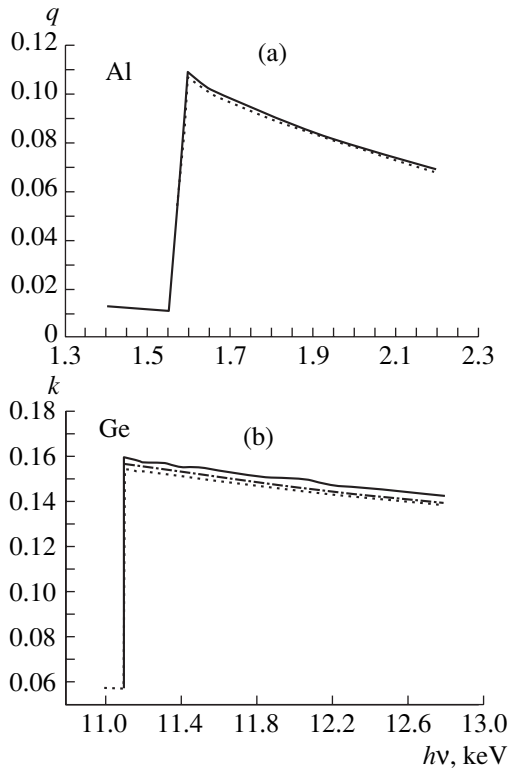


Fig. 3. Electron yield calculated analytically (dashed line) and by the Monte Carlo method (solid line) for (a) Al and (b) Ge. The dash-and-dot line in (b) shows the Monte Carlo simulation without taking into account fluorescent photons.

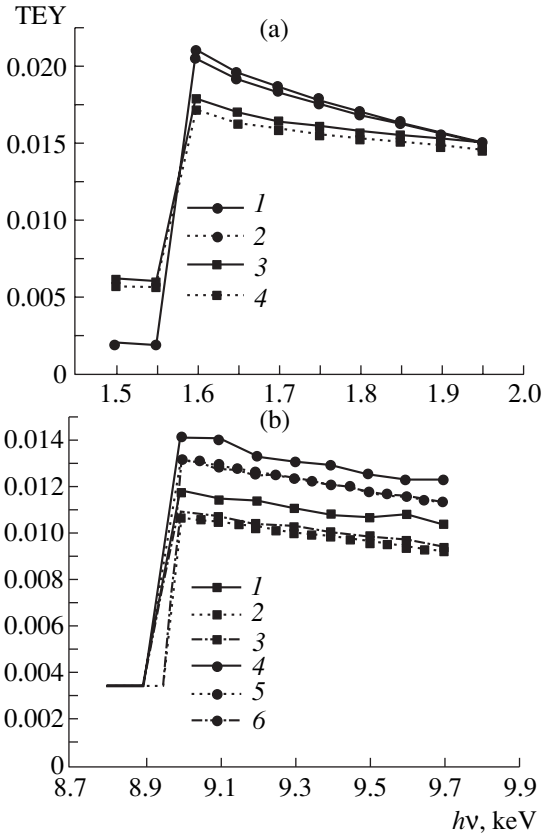


Fig. 4. TEY for the Al/Cu samples near the *K* edges of (a) Al and (b) Cu. (a) (*1*) Monte Carlo, $a = 0.6 \times 10^{-5}$ cm; (*2*) analysis, $a = 0.6 \times 10^{-5}$ cm; (*3*) Monte Carlo, $a = 0.1 \times 10^{-5}$ cm; and (*4*) analysis, $a = 0.1 \times 10^{-5}$ cm. (b) (*1*) Monte Carlo, $a = 0.6 \times 10^{-5}$ cm; (*2*) analysis, $a = 0.6 \times 10^{-5}$ cm; (*3*) Monte Carlo without fluorescent photons, $a = 0.6 \times 10^{-5}$ cm; (*4*) Monte Carlo, $a = 0.1 \times 10^{-5}$ cm; (*5*) analysis, $a = 0.1 \times 10^{-5}$ cm; (*6*) Monte Carlo without fluorescent photons, $a = 0.1 \times 10^{-5}$ cm.

any energy losses due to inelastic electron–target interaction are spent on the generation of the secondary electrons whose trajectories are then traced exactly in the same way as those of the primary electrons. With the contribution of the secondary electrons thus included, the TEY increases by several times as follows from the calculations. The energy of the secondary electrons is small compared with that of the primary electrons; therefore, the contribution of the former can be reduced substantially if electrons with an energy below some threshold (cutoff) value are omitted from consideration. Physically, this means the presence of an electrostatic barrier or a blocking potential near the surface. Taking into account the cutoff energy affects the escape function and the escape length of fast electrons (with initial energies above 2 keV) insignificantly, while for slow electrons, the presence of the cutoff energy is essential. For a cutoff energy of 100 keV, the discrepancy between the Monte Carlo simulation including the secondary electrons and the analytical results including the primary electrons alone is about 15%; for a cutoff energy of 250 keV, the discrepancy declines to 10%. In this way, one can estimate the contribution of relatively fast secondary electrons to the TEY.

Next, the TEY was evaluated for the Al/Cu and Cu/Al composites with the layer thickness varying from 10 to 60 nm. The thickness range was taken based on the data for the escape lengths for the most representative groups of Auger electrons (constituting the major portion of the primary electrons) escaping the homogeneous Al and Cu samples and the Al/Cu and Cu/Al composites.

The X-ray photon energy was varied from 1.50 to 1.95 keV near the *K* edge of Al and from 8.8 to 9.7 keV near that of Cu. The analytical and numerical (Monte Carlo) dependences of the TEY on the photon energy are demonstrated in Figs. 4 and 5 for an irradiation angle $\varphi = 30°$. For the Al/Cu composites (Fig. 4) at energies near the *K* edge of Al, the results are in good agreement. The discrepancy, which is about 3% for the 10-nm-thick layer and decreases in thicker layers, is associated with the inaccurate analytical approximation of the escape function for the substrate at large depths.

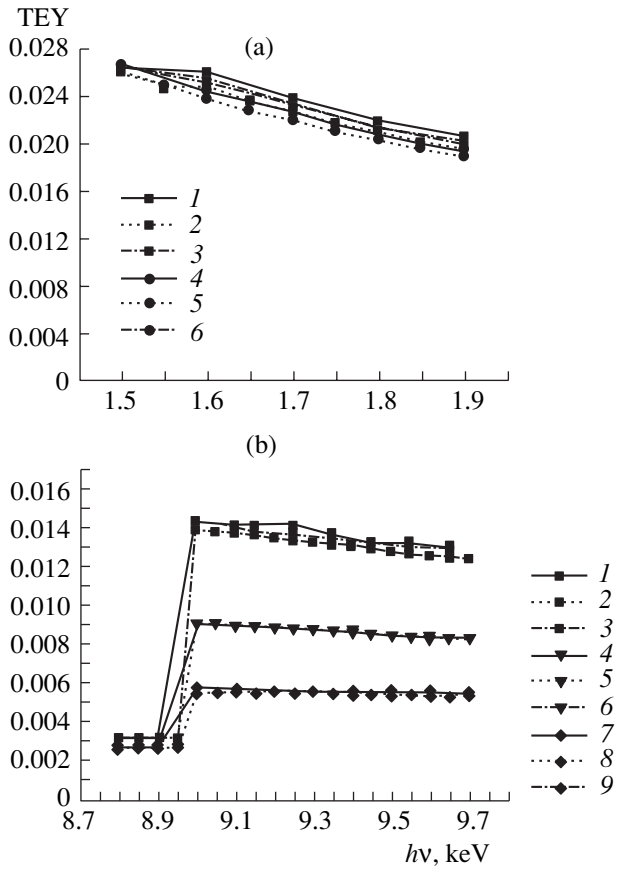


Fig. 5. TEY for the Cu/Al samples near the *K* edges of (a) Al and (b) Cu. (a) (*1*) Monte Carlo, $a = 0.1 \times 10^{-5}$ cm; (*2*) analysis, $a = 0.1 \times 10^{-5}$ cm; (*3*) Monte Carlo without fluorescent photons, $a = 0.1 \times 10^{-5}$ cm; (*4*) Monte Carlo, $a = 0.2 \times 10^{-5}$ cm; (*5*) analysis, $a = 0.2 \times 10^{-5}$ cm; and (*6*) Monte Carlo without fluorescent photons, $a = 0.2 \times 10^{-5}$ cm. (b) (*1*) Monte Carlo, $a = 1.6 \times 10^{-5}$ cm; (*2*) analysis, $a = 1.6 \times 10^{-5}$ cm; (*3*) Monte Carlo without fluorescent photons, $a = 1.6 \times 10^{-5}$ cm; (*4*) Monte Carlo, $a = 0.4 \times 10^{-5}$ cm; (*5*) analysis, $a = 0.4 \times 10^{-5}$ cm; (*6*) Monte Carlo without fluorescent photons, $a = 0.4 \times 10^{-5}$ cm; (*7*) Monte Carlo, $a = 0.1 \times 10^{-5}$ cm; (*8*) analysis, $a = 0.1 \times 10^{-5}$ cm; and (*9*) Monte Carlo without fluorescent photons, $a = 0.1 \times 10^{-5}$ cm.

At photon energies lower than the *K* edge of Al, the TEY depends largely on substrate electrons and decreases with increasing layer thickness. At energies above the *K* edge, the TEY grows with layer thickness.

Near the *K* edge of Cu, the analysis gives the TEY value roughly 10% lower than the numerical simulation. This difference is associated mainly with fluorescent photons contributing to the TEY. If the Monte Carlo simulation ignores fluorescent photons, the TEY value deviates from the analytical result by 3% for the 60-nm-thick layer and diminishes in thinner layers. As the aluminum layer thickens, the electron total emission from the copper drops. At photon energies lower than the *K* edge of copper, the TEY is almost independent of the aluminum thickness because of the minor contribution of electrons from the substrate.

For the Cu/Al samples (Fig. 5), the analytical and numerical results at photon energies near the *K* edge of aluminum differ by about 6%: ≈3% due to the contribution of fluorescent photons and ≈3% due to the inaccurate approximation of the escape function. A jump of the TEY at the *K* edge of aluminum is noticeable only at a copper layer thickness of 10 nm; for thicker layers, the probability of electrons escaping from the aluminum through the copper layer is negligible.

Near the *K* edge of copper, the results nearly coincide. The inaccuracy of the approximation of the escape function becomes essential with thick layers (160 nm), for which the TEYs for the homogeneous samples and composites are very close. Fluorescent photons contribute markedly in the case of very thin layers (most probably because of electrons generated in the substrate) and when the layer is very thick (probably because of the electron generation in the layer). In both cases, however, the contribution hardly exceeds the statistical error. As the copper layer thickness grows, the electron emission from the copper is enhanced.

CONCLUSION

The kinetic equation for medium-energy (several keV) electrons in composites comprising a layer of one material applied on a semi-infinite homogeneous substrate of another material was solved analytically in the diffusion approximation. From the solution obtained, analytical expressions for the escape function of electrons leaving the layer and the substrate were derived.

To verify these expressions and to refine the numerical values of parameters entering into them, the analytical results were compared with the results of numerical simulation by the Monte Carlo method. In the numerical simulation, we used the single scattering model. Elastic electron–material interaction was described in terms of the Mott differential cross section; inelastic interaction, in terms of the doubly differential inverse mean free path calculated within the dielectric formalism.

It was found that the escape lengths appearing in the analytical expressions for the escape function differ from the escape lengths in homogeneous samples and depend on the layer thickness and substrate composition. This discrepancy grows with decreasing layer thickness and increasing atomic number contrast. For the layer, the escape length can be determined by the numerical simulation of the escape function with the Monte Carlo method. When the analytical expressions for the escape function use the parameter values found by numerical simulation, they approximate the Monte Carlo results more closely.

Based on the analytical approximation of the escape function, we constructed a semianalytical algorithm for

calculating the TEY for the homogeneous samples and layer-on-substrate composites. According to this algorithm, the TEY is determined from the analytical expressions derived by solving the kinetic equation in the diffusion approximation. The numerical values of the parameters appearing in the analytical expressions (escape lengths and energy-dependent factors C) are found by the preliminary Monte Carlo simulation of the escape function. The TEY results calculated with such a procedure were compared with those obtained by the Monte Carlo method for samples of different composition. The effect of various processes on the final result was estimated. The data obtained with the analytical and numerical methods are in good agreement.

REFERENCES

1. *X-ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES*, Ed. by D. C. Koningsberger and R. Prins (Wiley, New York, 1991), Chemical Analysis, Vol. 92.
2. *Spectroscopy for Surface Science*, Ed. by R. J. H. Clark and R. E. Hester (Wiley, Chichester, 1998), Advances in Spectroscopy, Vol. 26.
3. *Lecture: New Brunch in the Non-destructive Solid Body Characterization "X-ray Induced Electron Emission at Absorption Edges. Ioffe Institute-95," 1995.*
4. L. A. Bakaleinikov, K. Ju. Pogrebitsky, E. A. Tropp, *et al.*, Nucleus **34**, 1 (1997).
5. L. A. Bakaleĭnikov, S. G. Konnikov, K. Yu. Pogrebitskiĭ, *et al.*, Zh. Tekh. Fiz. **64** (4), 9 (1994) [Tech. Phys. **39**, 354 (1994)].
6. L. A. Bakaleĭnikov, E. Yu. Flegontova, K. Yu. Pogrebitskiĭ, *et al.*, Zh. Tekh. Fiz. **71** (7), 14 (2001) [Tech. Phys. **46**, 796 (2001)].
7. L. A. Bakaleĭnikov and É. A. Tropp, Zh. Tekh. Fiz. **56**, 16 (1986) [Sov. Phys. Tech. Phys. **31**, 9 (1986)].
8. D. Pines, *Elementary Excitations in Solids* (Benjamin, New York, 1963; Mir, Moscow, 1965).
9. *Fundamental Studies on the Interactions of kV Electrons with Solids for Applications to Electron Spectroscopies*, Ph.D. Thesis (Osaka University, Osaka, 1990).
10. *Handbook of Optical Constants of Solids*, Ed. by E. D. Palik (Academic, New York, 1985).
11. E. Yu. Flegontova, L. A. Bakaleĭnikov, K. Yu. Pogrebitskiĭ, *et al.*, Zh. Tekh. Fiz. **70** (12), 6 (2000) [Tech. Phys. **45**, 1518 (2000)].
12. http://www.ioffe.rssi.ru/ES.
13. M. Ya. Amusia, *Atomic Photoeffect* (Plenum, New York, 1990).

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1193–1196. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 130–133.

BRIEF COMMUNICATIONS

Dynamics of a Two-Wave Packet in a Nonlinear Optical Fiber at Detuning from Phase Matching

I. O. Zolotovskiĭ and D. I. Sementsov

Ul'yanovsk State University, Ul'yanovsk, 432700 Russia

e-mail: sdi@sdi.ulsu.ru

Received September 5, 2001

Abstract—The effect of detuning from phase matching on the dynamics of a wave packet consisting of two unidirectional strongly interacting modes in a fiber with Kerr nonlinearity is considered. The effective parameters of dispersion and nonlinearity are analyzed for a wave packet that can be represented by a single partial pulse. © *2002 MAIK "Nauka/Interperiodica".*

(1) The propagation of coupled waves occupies a prominent place among the problems of nonlinear optics extensively discussed today. Coupled waves usually arise in tunnel-coupled or periodical-index optical fibers [1–3]. The unique dispersion properties of systems with strong linear wave coupling [4–6] motivates analysis of short-pulse propagation along such a fiber with allowance for nonlinear effects [7–9]. In [10], the propagation of short pulses was studied in a medium with Kerr nonlinearity under the condition of phase matching between two pulse-forming modes. In this paper, we study the effect of detuning from phase matching on the dynamics of a wave packet propagating in a periodical-index optical fiber with allowance for strong linear and nonlinear (cross-modulation) interactions between the modes.

(2) Let us consider a double-mode fiber periodical along its length. An optical pulse of duration τ_0 is applied to its input. Effective coupling between the pulse-forming unidirectional modes propagating in the fiber takes place in the phase matching range. Complete phase matching occurs at the center frequency ω_0 because of the fiber-index periodicity and pulse excitation. The condition for complete phase matching is the zero detuning δ at the center frequency:

$$\delta(\omega_0) = \beta_1(\omega_0) - \beta_2(\omega_0) - 2\pi/\Lambda = 0, \qquad (1)$$

where β_j is the wave number of jth mode (j = 1, 2) and Λ is the fiber inhomogeneity period along the axis of radiation propagation.

An equation for the temporal envelopes of two interacting modes written in the running time coordinates $\tau = t - z/u$, where u is the wave packet group velocity, is given by [10]

$$\frac{\partial A_j}{\partial z} + \frac{\xi_j}{v}\frac{\partial A_j}{\partial \tau} - i\frac{d_j}{2}\frac{\partial^2 A_j}{\partial \tau^2} = -i\sigma A_{3-j}\exp(i\xi_j\delta z) - iR(\gamma_s|A_j|^2 + \gamma_c|A_{3-j}|^2)A_j \qquad (2)$$

with allowance for the intermode detuning of group velocities, material dispersion, and nonlinear effects, such as phase self-modulation and cross-modulation. In Eq. (2), $\xi_j = (-1)^j$, $v^{-1} = (u_1 - u_2)/2u^2$ is the detuning of the mode inverse group velocities, $u_j = (\partial\beta_j/\partial\omega)_0^{-1}$ is the group velocity of the jth mode, $2u = u_1 + u_2$, $d_j = (\partial\beta_j^2/\partial\omega^2)_0$ is the material dispersion of the fiber, and R is the parameter of fiber nonlinearity. The coefficients of mode coupling σ and the parameters of phase self-modulation, γ_s, and cross-modulation, γ_c, are defined by the corresponding overlap integrals for the waveguide mode profile functions with regard for the optical inhomogeneity profile and the distribution of the modulation depth over the fiber [5, 6].

It should be emphasized that Eq. (2) applies to all systems with linearly coupled unidirectional waves. For example, the equations describing optical pulse propagation in tunnel-coupled light guides coincide with Eq. (2); in this case, however, the detuning parameter depends only on the difference between the mode propagation constants, $\delta = \beta_1 - \beta_2$, and the parameters σ, γ_s, and γ_c are given by formulas somewhat differing from those for a periodical fiber [1].

1063-7842/02/4709-1193$22.00

Equations (2) should be solved together with initial conditions for the mode temporal envelopes A_j that correlate with fiber excitation conditions. In the general form, the initial conditions are given by $A_2(\tau, 0) = \psi A_1(\tau, 0)$, where the parameter ψ governs the excitation of the fiber. For $\psi = \pm 1$, the excitation is symmetric or antisymmetric, respectively, while for $\psi = 0$ or $\psi^{-1} = 0$, we have single-mode excitation. For our case of strong mode coupling, the envelope of the corresponding mode can be represented as a combination of two partial pulses (PPs) [5, 6]:

$$\begin{aligned} A_1 &= a_1(\tau, z)\exp[i(q + \delta/2)z] + a_2(\tau, z) \\ &\quad \times \exp[-i(q - \delta/2)z], \\ A_2 &= \chi a_1(\tau, z)\exp[i(q - \delta/2)z] \\ &\quad - \chi^{-1} a_2(\tau, z)\exp[-i(q + \delta/2)z]. \end{aligned} \tag{3}$$

Here, a_f are the PP amplitudes slowly varying with the z coordinate, $q \equiv (\sigma^2 + \delta^2/4)^{1/2}$, and the parameter

$$\chi = \frac{(2q + \delta)A_{20} - 2\sigma A_{10}}{(2q - \delta)A_{10} - 2\sigma A_{20}} = -\frac{(2q + \delta)A_{10} + 2\sigma A_{20}}{(2q - \delta)A_{20} - 2\sigma A_{10}}$$

depends on the initial conditions of fiber excitation.

In the case of complete phase matching, $\delta = 0$, $q = |\sigma|$, and $\chi = -1$.

Thus, the pulse formed by two interacting modes is the superposition of partial pulses with the amplitudes satisfying the equation

$$\begin{aligned} &\frac{\partial a_f}{\partial z} - \frac{\xi_f \delta}{2qv}\frac{\partial a_f}{\partial \tau} - \frac{iD_f}{2}\frac{\partial^2 a_f}{\partial \tau^2} \\ &+ iR(\gamma_{sf}|a_f|^2 + \gamma_{cf}|a_{3-f}|^2)a_f = 0 \end{aligned} \tag{4}$$

in view of (2) and (3). Here, $f = 1, 2$; $\xi_f = (-1)^f$; and the effective parameters of phase self-modulation and cross-modulation, as well as the PP effective dispersion, are given, respectively, by

$$\gamma_{sf} = \frac{(\gamma_s + \gamma_c)}{2}\left(1 + \chi^{-2\xi_f} + \frac{\xi_f \delta}{2q}\frac{(\gamma_s - \gamma_c)}{(\gamma_s + \gamma_c)}(1 - \chi^{-2\xi_f})\right),$$

$$\gamma_{cf} = \frac{(\gamma_s - \gamma_c)}{2} \tag{5}$$

$$\times\left(\frac{(3 + \chi^{2\xi_f})\gamma_s - (1 - \chi^{2\xi_f})\gamma_c}{\gamma_s - \gamma_c} + \frac{\xi_f \delta}{2q}(3 - \chi^{2\xi_f})\right),$$

$$D_f = \frac{d_1 + d_2}{2} + \frac{\xi_f}{v^2 q}\left(1 - \frac{\delta v^2 (d_1 - d_2)}{2}\right). \tag{6}$$

In view of (4), the initial conditions for the PP amplitudes have the form

$$a_f(\tau, 0) \equiv a_{f0} = \frac{1}{2}\left[A_{10} + \xi_f\left(\frac{\delta}{2q}A_{10} + \frac{\sigma}{q}A_{20}\right)\right]. \tag{7}$$

The "degenerate situation," where the behavior of the whole wave packet can be described by the dynamics of one of the PPs, is of most interest for deriving analytical solutions to Eq. (4) and their experimental realizations. For such a situation, it is possible to find optimal dispersion parameters of the pulse propagating in a fiber. In the case of phase matching of the modes ($\delta = 0$), degeneration takes place for symmetric ($\psi = 1$) or antisymmetric ($\psi = -1$) excitations, when $a_{10} = 0$, $a_{20} \neq 0$ or $a_{20} = 0$ and $a_{10} \neq 0$, respectively. Here, one of the PP amplitudes a_f equals zero at the initial time instant and remains the same during the pulse propagation.

(3) In the general case $\delta \neq 0$, i.e., in the case of detuning from phase matching, degeneration for the asymmetric excitation is also possible, according to (7), if $a_1 = 0$ at $\psi = -(2q - \delta)/2\sigma$ or $a_2 = 0$ at $\psi = (2q + \delta)/2\sigma$. In this case, the system of equations (4) degenerates into the single nonlinear Schrödinger equation [10]

$$\frac{\partial a_f}{\partial z} - \frac{iD_f}{2}\frac{\partial^2 a_f}{\partial^2 \tau_f^2} + iR\gamma_{sf}|a_f|^2 a_f = 0, \tag{8}$$

where the running time $\tau_f = t - z/u_f$ is related to the corresponding PP and the PP group velocity is $u_f^{-1} = u^{-1} - \xi_f \delta/2qv$.

Equation (8) describes the pulse dynamics in a cubically nonlinear medium with an effective dispersion D_f and an effective nonlinearity $R\gamma_{sf}$. A distinctive feature of the wave packet propagation in such a medium (see Eq. (8)) is self-action, which may both broaden and compress the wave packet. Another feature of the process is the possibility of forming stable wave packets, Schrödinger solitons, appearing under the combined action of the nonlinearity of the medium and dispersion effects [11]. In our case of strong linear coupling between the modes forming a wave packet, the dispersive properties depend on the effective dispersion D_f of a partial pulse due to the material dispersion, mode coupling, and detuning from phase matching. If the effective dispersion is anomalous ($D_f < 0$) and a waveguide has focusing properties ($R\gamma_{sf} > 0$), Eq. (8) has a solution specifying so-called "bright" solitons of hyperbolic secant shape. In this case, the solution of Eq. (8) for the PP amplitude is

$$a_f(\tau, z) = a_{f0}\,\mathrm{sech}(\tau/\tau_s)\exp(-i\Gamma z), \tag{9}$$

where the phase, duration, and initial amplitude of the pulse (soliton) are related as

$$2\Gamma = R\gamma_{sf}a_{f0}^2 = |D_f|/\tau_{sol}^2. \quad (10)$$

It follows from Eq. (10) that the soliton duration depends strongly on the effective dispersion and nonlinearity values:

$$\tau_{sol} = (|D_f|/R\gamma_{sf}a_{f0}^2)^{1/2} = (|D_f|\tau_0/R_fW_0)^{1/2}. \quad (11)$$

Here, $W_0 = (|A_{10}|^2 + |A_{20}|^2)\tau_0$ is the density of the energy delivered to the fiber and $R_f = R\gamma_f$ is the parameter of effective nonlinearity, where

$$\gamma_f = \frac{\gamma_s + \gamma_c}{2}\left(1 + \frac{\delta^2}{4\sigma^2 + \delta^2}\frac{(\gamma_s - \gamma_c)}{(\gamma_s + \gamma_c)}\right). \quad (12)$$

It is worth noting that in the degenerate case, where the behavior of the entire wave packet is governed by one PP, the parameter of effective nonlinearity does not depend on the excitation type, and the energy density of a soliton-like pulse passing through the fiber is $W_{solf} = |D_f|/R_f\tau_{sol}$. If the energy density W_0 applied to the fiber is close to W_{solf}, the soliton regime of pulse propagation sets in. For $W_0 < W_{solf}$, the pulse diffuses; for $W_0 > W_{solf}$ it shrinks. In the latter case, the pulse self-compression can be approximated by the generalizing relationship [12]

$$\tau_0/\tau_{min} \cong 4\sqrt{\tau_0 R_f W_0/|D_f|}, \quad (13)$$

where τ_{min} is the minimum pulse duration.

The effective parameters of dispersion and nonlinearity depend appreciably on the detuning value, mode coupling, and excitation type, providing conditions for effectively controlling the compression, which to a great extent specifies the pulse dynamics in the fiber. The effective parameters of dispersion and nonlinearity vs. the normalized detuning from phase matching $\delta/|\sigma|$ for partial pulses with $f = 1$ (*a*) and $f = 2$ (*b*) and for the fiber parameters $v^2|\sigma| = 10^{26}$ m/s^2, $d_1 = 10^{-26}$ s^2/m, and $d_2 = (-3, -2, -1, 1, 2, 3)10^{-26}$ s^2/m (curves *1*–*6*) are given in Fig. 1. The monotonical increase in D_1 is accompanied by the monotonical decrease in D_2 (and conversely) at $d_1 \neq d_2$. For $d_1 = d_2$, the weak dependence of D_f on the detuning is observed. For $\delta = -2\xi_f|\sigma|/\sqrt{3}$, the d_2-dependence of D_f disappears: $D_f = d_1 + \xi_f/v^2q$; similarly, for $\delta = 2\varepsilon_f|\sigma|/\sqrt{3}$, D_f does not depend on d_1. Of greatest importance is the possibility of reaching extremely small values of the effective dispersion D_f (less than $|10^{-27}|$ s^2/m) by varying the detuning, which is a function of the phase matching δ. The possibility to attain small and, at the same time, negative effective dispersion over a wide range of operating frequencies,

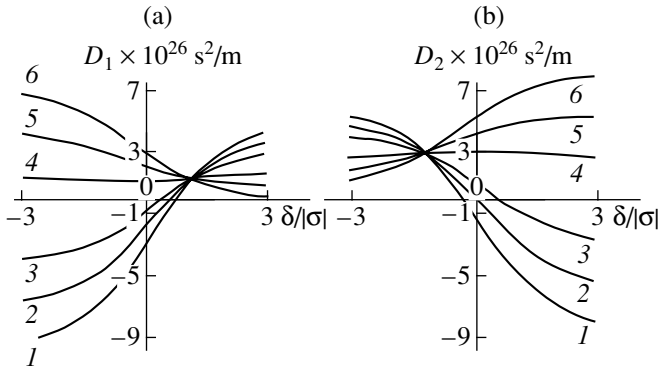


Fig. 1. Effective dispersion D_f vs. the normalized detuning $\delta/|\sigma|$ for partial pulses with $f = 1$ (a) and 2 (b).

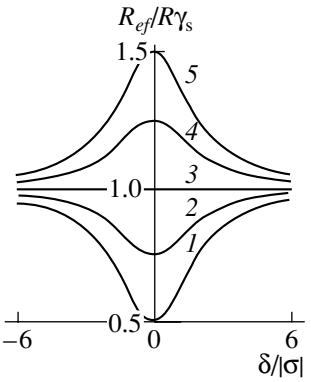


Fig. 2. Normalized parameter of the effective nonlinearity $R_{ef}/R\gamma_s$ vs. the normalized detuning $\delta/|\sigma|$.

$\omega_0 = 0.8$–4.0×10^{15} s^{-1}, is of great value for developing compact laser-pulse compressors based on the optical fibers considered.

The dependences of the normalized parameter of the effective nonlinearity $R_f/R\gamma_s$ on the detuning $\delta/|\sigma|$ for $\gamma_c/\gamma_s = 0$, 0.5, 1, 1.5, and 2 (curves *1*–*5*) are shown in Fig. 2. The effective nonlinearity is the same for both PPs and does not depend on the detuning sign. For $\delta = 0$, we have the extreme value of nonlinearity $R_f \cong (\gamma_s + \gamma_c)R/2$; in the limit of high values of the detuning, $|\delta| \Rightarrow \infty$, $R \Rightarrow R\gamma_s$. It is remarkable that the effective nonlinearity is detuning-independent for $\gamma_c = \gamma_s$. The maximum change in the nonlinearity parameter is $R|\gamma_s - \gamma_c|/2$.

The essential effect of radiation input conditions on the radiation dynamics in the fibers makes them promising for all-optical logic elements. It should be stressed that coupled waves can be generated not only in periodical-index media but also in tunnel-coupled fibers, gyrotropic birefringent media, etc. A periodical-index fiber can be considered as a model object, and associated results should be viewed as general and applicable to almost any system supporting copropagating waves with strong linear coupling.

REFERENCES

1. A. A. Maĭer, Usp. Fiz. Nauk **165**, 1037 (1995) [Phys. Usp. **38**, 991 (1995)].
2. S. A. Vasil'ev, E. M. Dianov, A. S. Kurkov, *et al.*, Kvantovaya Élektron. (Moscow) **24** (10), 151 (1997).
3. S. A. Vasil'ev, E. M. Dianov, D. S. Starodubov, *et al.*, Kvantovaya Élektron. (Moscow) **24** (10), 160 (1997).
4. V. A. Vysloukh and L. P. Gevorkyan, Izv. Akad. Nauk SSSR, Ser. Fiz. **55** (2), 323 (1991).
5. I. O. Zolotovskiĭ and D. I. Sementsov, Opt. Spektrosk. **86**, 824 (1999) [Opt. Spectrosc. **86**, 737 (1999)].
6. I. O. Zolotovskiĭ and D. I. Sementsov, Kvantovaya Élektron. (Moscow) **27**, 273 (1999).
7. A. I. Maĭmistov, Kvantovaya Élektron. (Moscow) **18**, 758 (1991).
8. F. Kh. Abdullaev, R. M. Abramov, V. I. Goncharov, and S. A. Darmanyan, Zh. Tekh. Fiz. **64** (9), 102 (1994) [Tech. Phys. **39**, 916 (1994)].
9. V. A. Malomed, P. L. Skinner, P. L. Chu, and G. D. Peng, Phys. Rev. E **53**, 4084 (1996).
10. I. O. Zolotovskiĭ and D. I. Sementsov, Opt. Spektrosk. **88**, 620 (2000) [Opt. Spectrosc. **88**, 560 (2000)].
11. S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *The Optics of Femtosecond Laser Pulses* (Nauka, Moscow, 1988).
12. E. M. Dianov, Z. S. Nikonova, A. M. Prokhorov, and V. N. Serkin, Pis'ma Zh. Tekh. Fiz. **12**, 756 (1986) [Sov. Tech. Phys. Lett. **12**, 311 (1986)].

Translated by M. Astrov

Technical Physics, Vol. 47, No. 9, 2002, pp. 1197–1198. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 134–135.

BRIEF COMMUNICATIONS

Radiative and Photoelectric Properties of $CuInS_2$ Single Crystals

V. A. Ivanov, I. A. Viktorov, and V. F. Gremenok

Institute of Solid-State and Semiconductor Physics, Belarussian Academy of Sciences, ul. Brovki 17, Minsk, 220072 Belarus

e-mail: gremenok@ifttp.bas-net.by

Received November 20, 2001

Abstract—The photoluminescence and photocurrent spectra of $CuInS_2$ single crystals grown by the Bridgman method are studied at temperatures of 80 and 300 K. The photosensitivity spectrum is observed in the short-wave photoluminescence band. From the watt–ampere characteristics of photoconductivity, a linear mechanism of recombination of minority charge carriers is established for an illumination level of up to 100 mW/cm^2 in the temperature range of 80–300K. © *2002 MAIK "Nauka/Interperiodica".*

The $CuInS_2$ semiconductor belongs to I–III–VI_2 compounds and has the chalcopyrite structure [1]. This direct-gap semiconductor has a bandgap E_g = 1.55 eV and a high optical absorption coefficient ($\alpha > 10^5$ cm^{-1}), as well as features intense radiative recombination. When combined, these properties make it a promising material for optoelectronic devices operating in the near-IR and visible spectral ranges, for example, for solar cells [2].

In this study, we report the radiative and photoelectric properties of undoped $CuInS_2$ single crystals grown by the Bridgman method. They had a *p*-type resistivity ρ = 0.1–1.0 Ω cm. The conductivity type was determined by the thermoelectric method. Samples to be studied were cleaved out of an as-grown ingot, and the cleavage surface was shiny. This crystal is known [3] to cleave over the crystallographic plane (112); therefore, we assumed this plane to be the cleavage plane in our case. Graphite contacts were deposited on the as-cleaved surface to perform electrical measurements. As follows from the *I–V* characteristics, these contacts are ohmic at temperatures between 80 and 300 K. The crystals grown had the chalcopyrite structure with parameters coinciding with those presented in [4, 5] to within an experimental error.

The radiative and photoelectric properties of the samples were studied with the setup described in [6]. A He–Ne laser (λ_0 = 0.633 μm) with a 0.1- to 1.0-W/cm^2-power density was used as an excitation source. A nitrogen-vapor-cooled FÉU-112 photoelectric multiplier with spectral characteristic similar to that of an S-1 photocathode served as a photodetector.

Figure 1 presents the photoluminescence spectra of the $CuInS_2$ single crystals. The spectra are seen to have three peaks at E_1 = 1.52 eV, E_2 = 1.43 eV, and E_3 = 1.11 eV at T = 300 K and at E_1' = 1.52 eV, E_2' = 1.48 eV, and E_3' = 1.11 eV at T = 80 K. The peak at E_1 can be attributed to optical transitions between shallow donors (D_1) and the valence band. The difference in E_1 values at 300 and 80 K stems from the temperature variation of the bandgap E_g in $CuInS_2$ crystals. A donor D_1 with an ionization energy ΔE_{D1} = 35 meV is known [2, 7] to be a sulfur vacancy (V_S). The peak at E_2 is associated with D_1–A_1 optical transitions, where an acceptor level A_1 with an ionization energy ΔE_{A1} = 100 meV is [2] a copper vacancy (V_{Cu}). At a temperature T = 80 K, the band peaked at E_2 is replaced by the band with a maximum at E_2'. The new band results from optical transitions from a donor level D_2 with an ionization energy ΔE_{D2} = 72 meV to the valence band. This level is also associated with sulfur vacancies, which form [8] two energy levels in the $CuInS_2$ bandgap. Both were observed in our optical absorption spectra (omitted in the present discussion). The intense long-wave temperature-independent band with a peak at E_3 in the photoluminescence spectrum can be attributed to D_1–A_2 tran-

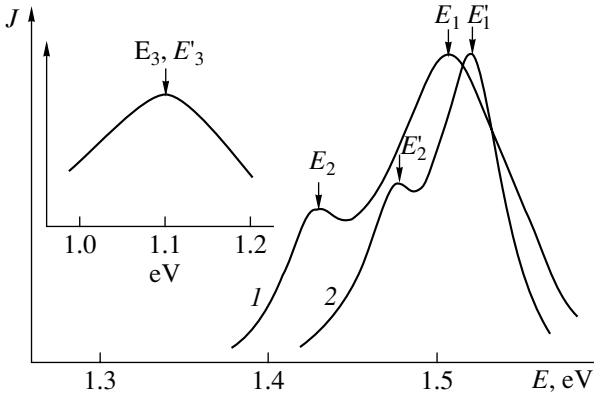


Fig. 1. Photoluminescence spectra of the $CuInS_2$ single crystals at (*1*) 300 and (*2*) 80 K.

1063-7842/02/4709-1197$22.00

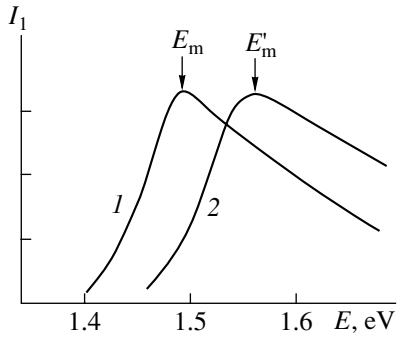


Fig. 2. Photocurrent spectra of the $CuInS_2$ single crystals at (*1*) 300 and (*2*) 80 K.

sitions, where A_2 is a deep acceptor level (ionization energy ΔE_{A2} = 370 meV) due to either growth-related oxygen interstitials or copper vacancies (V_{Cu}) [8].

Photocurrent spectra of the crystals were also taken at 300 and 80 K (Fig. 2). They take into consideration the spectral distribution of the radiation source, which was a KGM halogen lamp with a color temperature T = 1400 K in the operating range. A single maximum with steep short-wave and smooth long-wave edges is observed. The positions of the maximum are E_m = 1.49 eV and E'_m = 1.55 eV at 300 and 80 K. The steepness of the short-wave edge displays a weak temperature dependence. Upon cooling the sample, the short-wave edge of the band shifts toward higher energies at a rate of 2.27×10^{-3} eV/K, which agrees with the rate of temperature variation of the $CuInS_2$ bandgap [2]. The smooth long-wave edge of the photocurrent spectra evidences a low rate of surface recombination of nonequilibrium charge.

In order to determine the mechanism behind nonequilibrium carrier recombination, we studied the illumination dependences of the photocurrent $I_1 = I_2 - I_3$ (I_2 is the current across the sample illuminated and I_3 is the dark current) at 300 and 80 K. The illumination was measured by an IMO-2N optical power meter. The photocurrent value appeared to be a linear function of the illumination for intensities of up to 100 mW/cm^2 for both 300 and 80 K. This suggests a linear mechanism of nonequilibrium carrier recombination in $CuInS_2$ crystals; i.e., the product $\mu\tau$ (μ is the carrier mobility and τ is the lifetime) is independent of the charge carrier concentration [9].

Thus, we found that the photosensitivity spectrum of $CuInS_2$ crystals lies in the short-wave part of the photoluminescence band. The smooth long-wave photocurrent edge indicates the slow surface recombination of minority charge carriers on the as-cleaved $CuInS_2$ single-crystal surface. Based on the watt–ampere characteristics of photoconductivity, we established the linear character of minority carrier recombination in the temperature range from 80 to 300 K at an illumination level of up to 100 mW/cm^2.

ACKNOWLEDGMENTS

This work was supported by the Belarussian Foundation for Basic Research.

REFERENCES

1. J. L. Shay and J. H. Wernick, *Ternary Chalcopyrite Semiconductors* (New York, 1975).
2. *Current Topics in Photovoltaics*, Ed. by T. Coutts and J. Meakin (Academic, London, 1985; Mir, Moscow, 1988).
3. S. Wagner, Top. Appl. Phys. **17**, 171 (1997).
4. I. V. Bodnar, I. T. Bodnar, and A. A. Vaipolin, Cryst. Res. Technol. **19**, 1533 (1984).
5. I. V. Bondar', Izv. Akad. Nauk SSSR, Neorg. Mater. **22**, 1369 (1991).
6. N. N. Koren' and V. A. Ivanov, Prib. Tekh. Éksp., 161 (1981).
7. G. Masse, N. Lahlou, and C. Butti, J. Phys. Chem. Solids **12**, 449 (1981).
8. G. Masse and E. Redjal, J. Phys. Chem. Solids **47**, 99 (1986).
9. M. Ya. Bagirov, *Semiconductor Detectors of Radiation* (Baku, 1983).

Translated by A. Sidorova

Technical Physics, Vol. 47, No. 9, 2002, pp. 1199–1201. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 136–139.

BRIEF COMMUNICATIONS

Magnetization Distribution in Submicrometer Iron–Nickel Antidot Arrays Patterned by a Focused Ion Beam

A. Yu. Toporov

Natural-Science Research Center, General Physics Institute, Russian Academy of Sciences, Moscow, 117991 Russia

e-mail: topor@kapella.gpi.ru

Received January 29, 2002

Abstract—Antidots of size 0.5 μm are prepared by patterning iron–nickel films with a focused ion beam. The magnetization distribution in antidot arrays is examined with Lorentz transmission electron microscopy. It is shown that one side of the array makes an angle of about 20° with the easy magnetic axis of the film. Magnetization reversal in the direction close to the easy magnetic axis starts with domain nucleation at the antidot edges that are perpendicular to the applied field and adjacent to the unpatterned region of the film, and propagates as the domain walls move. Magnetization reversal in the direction close to the hard magnetic axis starts with magnetization rotation outside the patterned region at the antidot edges and propagates as the domain walls execute a complicated motion. It is demonstrated that some areas between the edges of adjacent antidots can carry information bits. Results obtained are explained in terms of competition between the demagnetizing energy, energy of internal anisotropy, and misorientation effect. The feasibility of such structures as high-density storage elements is discussed. © *2002 MAIK "Nauka/Interperiodica".*

One of the most important problems in today's magnetic technologies is to raise the density of magnetic disks. The goal can be achieved by decreasing the grain size of a magnetic material or by reducing the energy exchange between grains. In both cases, however, great difficulties associated with the thermal instability of the grain magnetic moment arise. A possible way of solving this problem is to lithographically split the magnetic material into islands with a nonmagnetic medium between them so that one island carries one data bit. Such structures are referred to as magnetic dots [1]. Recent publications describe an alternative configuration, namely, nonmagnetic dots surrounded by a magnetic material. By analogy, they have been called magnetic antidots. In [2–5], micromagnetic studies of magnetic antidot arrays with magnetic-force microscopy and scanning magnetooptic Kerr microscopy are reported. In addition, the behavior of periodic antidot arrays and writing processes in these media were simulated [6]. However, results obtained in the works cited need experimental verification. Later, magnetic antidot arrays patterned by a focused ion beam (FIB) on continuous films were investigated with Lorentz transmission electron microscopy [7]. In this work, we study the magnetization distribution in submicrometer antidot arrays prepared by the FIB patterning of $Ni_{80}Fe_{20}$ continuous films by means of Lorentz TEM with a magnetic film applied *in situ*.

Antidots were patterned in 20-nm-thick $Ni_{80}Fe_{20}$ polycrystalline films deposited on carbon-covered mica substrates by magnetron sputtering in argon. The deposition rate was 3.3 nm/min, and the pressure in the vacuum chamber was 10^{-6} mbar at room temperature and 6×10^{-3} mbar during the deposition process. As-deposited films were floated off in water and then placed on 200×200-μm square-mesh copper grids.

Square arrays were patterned in the films with a 200TEM FIB system (FEI Co., USA). Ga ions accelerated by a voltage of 30 kV were used as a milling species. The ion beam and the spot size were 0.5 nA and 10 nm, respectively. The antidot size and the size-to-spacing ratio were 0.5 μm and 1 : 2, respectively.

The energy-dispersive X-ray analysis of patterned areas was carried out by recording characteristic X-ray spectra from the samples as a result of electron bombardment in a JEOL 3000F electron microscope with a field-emission electron gun (FEG).

The Lorentz TEM study of the domain structure of the antidot arrays was carried out with a JEOL 4000EX electron microscope modified for magnetic films [8]. A magnetic field of strength H up to ±400 Oe can be applied *in situ* in the plane of the film parallel to the edges of antidots. Before the experiments, the film was gradually magnetized to saturation in both the forward and reverse directions. Then, the image of the domain structure appearing in the array in both the state with remanent magnetization and the biased state was photographed.

The TEM studies showed that the antidots were not square in shape because of the nonuniform energy distribution over the ion beam cross section. As followed from the energy-dispersive X-ray analysis, the patterned areas were free of the substrate material ele-

1063-7842/02/4709-1199$22.00

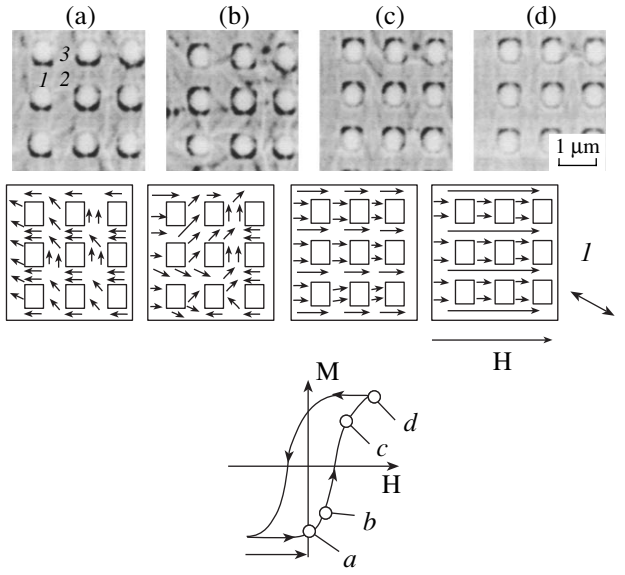


Fig. 1. Lorentz TEM images of the antidot array that are obtained upon magnetization reversal in direction *1* close to the EMA. The magnetization distribution is schematically shown in the diagrams below. H = 0 (a), 19 (b), 34 (c), and 57 Oe (d).

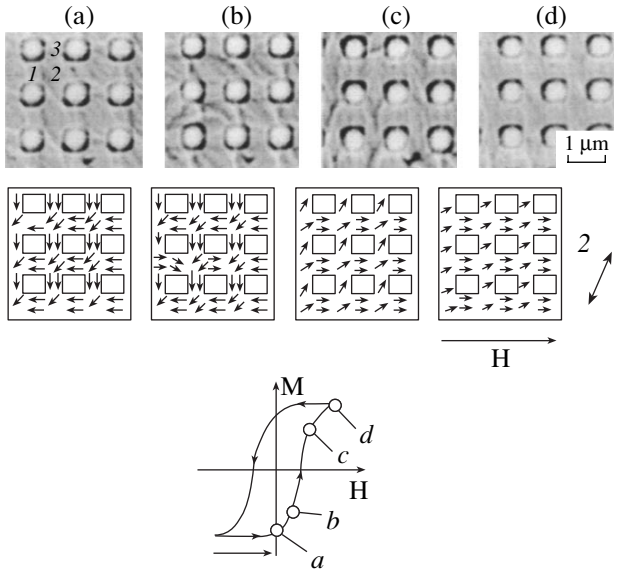


Fig. 2. Lorentz TEM images of the antidot array that are obtained upon magnetization reversal in direction *2* close to the HMA. The magnetization distribution is schematically shown in the diagrams below. (a–d) The same as in Fig. 1.

ments, and their edges were sharp. The Lorentz TEM investigation demonstrated that the arrays make an angle of about 20° with the easy magnetic axis (EMA) of the $Ni_{80}Fe_{20}$ film. The EMA was defined as the direction running normally to the magnetization ripple. Figure 1a shows the magnetization distribution over the antidot arrays in the state with remanent magnetization before the film is magnetized in the direction close to the EMA. Schematically, the magnetization distribution is indicated by arrows in the diagram depicted below. Three regions with various mean directions of magnetization can be distinguished. In region *1*, the magnetization is oriented horizontally because of the demagnetizing effect at the edges of the antidots. In region *2*, the misorientation between the EMA and edges of the antidots leads to the diagonal direction of magnetization. In region *3*, the demagnetizing energy is roughly equal to the energy of internal anisotropy. Because of this, the magnetization orientation in these regions depends on local variations of the anisotropy variance and film inhomogeneity. Magnetization reversal in the direction close to the easy magnetic axis starts with the nucleation of inversely magnetized domains at the antidot edges that are perpendicular to the applied field and adjacent to the unpatterned region outside the array. In a field of about 14 Oe, these domains begin to penetrate into the array along rows aligned with the applied field (regions *1* and *2*). The magnetization mean direction in these domains coincides with the EMA because of the misorientation effect. Individual domain walls are pinned at the edges of the antidots, forming a long domain wall. Meanwhile, at a field near 16 Oe, the motion of the domain walls starts to switch regions *3*. Figure 1b shows the magnetization distribution over the array in a field of 19 Oe. When the applied field exceeds 21 Oe, the magnetizations in regions *1* and *2* are aligned with the field. However, this field is insufficient to overcome the demagnetizing field in regions *3*, as follows from Fig. 1c. Some of the domains remain pinned between the corners or edges of adjacent antidots, since the applied field is still insufficient to overcome the energy barriers at the antidot edges. As the field rises, the magnetization vectors rotate toward its direction (Fig. 1d). When the field diminishes, the magnetic field returns to the state with remanent magnetization. With the field reversed, the magnetization switching proceeds in the reverse direction.

In the direction close to the hard magnetic axis (HMA), magnetization reversal occurs in a different way. Figure 2a shows the magnetization distribution in the antidot array in the state with remanent magnetization before the magnetic field is applied *in situ*. A noticeable feature here is the existence of regions where the magnetization is distributed in a similar way in the absence of the field. In these regions (denoted *1*), the magnetization vectors run horizontally because of the demagnetizing effect. In regions *2* and *3*, the effect of misorientation between the EMA and the antidot edges and the demagnetizing effect cause the magnetization vectors to run diagonally and vertically, respectively. Magnetization reversal starts largely with magnetization rotation at the antidot edges outside the array and becomes detectable at an applied field of 65 Oe. At 9 Oe, the magnetization vector rotation begins to switch the magnetizations in regions *2*, as indicated by a faint change in the magnetization ripple orientation with the

domain walls remaining immobile. As the field strength rises, inversely magnetized domains formed at the antidot edges normal to the field grow, merge together, and penetrate into the array, partially changing the magnetization distribution in regions *1* (lower left-hand square in Fig. 2b). Simultaneously, at a field of about 19 Oe, magnetization reversal starts in regions *2*. In these regions, the reversal takes place through a sudden jump of the domain walls, with the vector of domain motion making an angle of about 45° with the antidot edges. This process is completed near 25 Oe. The range of switching fields is associated with the magnetic field nonuniformity. Figure 2c depicts the array subjected to a field of 34 Oe. In regions *1*, the demagnetizing field facilitates the magnetization orientation along the applied field. In regions *3*, the magnetization direction depends on the superposition of the applied field, the effective field of internal anisotropy, and the demagnetizing field. Finally, in regions *2*, where the effect of the demagnetizing field at the antidot edges is less significant, the directions of the magnetization vectors are close to that of the applied field. As for magnetization switching in the direction close to the EMA, the array contains domains pinned between the edges and corners of the antidots. The reason why such domains exist has been discussed earlier. At higher applied fields, the magnetization vectors rotate toward the applied field direction (Fig. 2d). Then, the applied field is reversed, and the process goes in the opposite direction. It should be noted that when the array is switched in the direction close to the HMA, the magnetization distribution in all regions *1* in the state with remanent magnetization have the same reproducible (from one hysteresis loop to another) configuration.

Thus, with Lorentz TEM, we studied the magnetization distribution in submicrometer antidot arrays prepared by patterning a continuous $Ni_{80}Fe_{20}$ film with an FIB. It was shown that magnetization reversal in the direction close to the EMA starts with the nucleation of inversely magnetized domain at the antidot edges (outside the array) that are perpendicular to the applied field and propagates through the ordinary motion of the domain walls. Magnetization reversal in the direction close to the HMA is initiated largely by magnetization rotation and propagates as the domain walls execute a complicated motion. In the array, there are regions between the edges of adjacent antidots where the magnetization distribution is reproduced from one hysteresis loop to another. Although these regions are not domains in the strict sense, they can be viewed as magnetic memory elements. However, the size of the antidots considered in this article is relatively large; therefore, they are inappropriate for elements of super-high-density memories. It is hoped that an array of smaller antidots with a size-to-spacing ratio and a magnetic material chosen optimally will be promising as a medium for super-high-density memory.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (project no. 01-02-17753) and a NATO/Royal Society Fellowship.

The author thanks R.M. Langford for assistance with the FIB patterning and also R.S. Doole for the assistance with the Lorentz TEM measurements.

REFERENCES

1. R. L. White, R. M. H. New, and R. F. W. Pease, IEEE Trans. Magn. **33**, 990 (1997).
2. R. P. Cowborn, A. O. Adeyeye, and J. A. C. Bland, Appl. Phys. Lett. **70**, 2309 (1997).
3. R. P. Cowborn, A. O. Adeyeye, and J. A. C. Bland, J. Magn. Magn. Mater. **173**, 193 (1997).
4. Y. Otani, S. G. Kim, T. Kohda, *et al.*, IEEE Trans. Magn. **34**, 1090 (1998).
5. C. T. Yu, H. Jiang, L. Shen, *et al.*, J. Appl. Phys. **87**, 6322 (2000).
6. L. Torres, L. Lopez-Diaz, O. Alejos, *et al.*, Physica B (Amsterdam) **275**, 59 (2000).
7. A. Yu. Toporov, R. M. Langford, and A. K. Petford-Long, Appl. Phys. Lett. **77**, 3063 (2000).
8. X. Portier and A. K. Petford-Long, J. Phys. D **32**, R89 (1999).

Translated by V. Isaakyan

Technical Physics, Vol. 47, No. 9, 2002, pp. 1202–1204. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 72, No. 9, 2002, pp. 140–142.

BRIEF COMMUNICATIONS

The Stability of a Group Cathode Spot on the Surface of Hot-Rolled Steels

S. L. Pozharov*, A. M. Mirkarimov*, and K. Takeda**

** Institute of Electronics, Academy of Sciences of Uzbekistan, Tashkent, 700143 Uzbekistan*

e-mail: <ariel@uzsci.net>

*** Department of Machine Intelligence and System Engineering, Akita Prefectural University, Honjyo, Akita, 015-0055, Japan*

Received March 13, 2002

Abstract—A general kinetic equation is derived for elementary cathode spots (ECSs) constituting a group cathode spot (GCS). The equation makes it possible to evaluate the kinetic constants of the spots if certain experimental data are available. ECS kinetics features are established. An explanation for the possibility of the stable state of a GCS is proposed. © *2002 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Investigation into GCSs in a vacuum arc discharge on scale-coated surfaces [1, 2] aimed at elaborating surface-treatment technology in such discharges [3] has revealed the following. First, a GCS is stable with characteristic relaxation times of its parameters in the range 100–200 μs. Second, a stable GCS has a configuration where ECSs are located along the GCS region boundary, i.e., along the metal–scale interface. Such a configuration has been called one-dimensional. Third, it has been established that the GCS configuration and, hence, the microrelief of the surface vary with the surface condition.

In this study, an effort is made to find a simple explanation for the GCS stability, since this problem is of great practical importance. To this end, we use a quasi-particle approximation for describing the ECS kinetics inside a GCS.

THE KINETICS OF ELEMENTARY CATHODE SPOTS

As state (phase) variables of a GCS, one can choose the number of ECSs in a GCS and any GCS size, for example, its perimeter. In describing the kinetics of ECSs on the cathode surface, they can be represented as quasi-particles in the two-dimensional space having a finite lifetime and being capable of generating the same new quasi-particles. Here, we use the same ideas and definitions regarding ECSs as in [2].

In the general form, the kinetic equations can be written as follows:

$$\frac{dN}{dt} = \sum_{i=1}^{m} k_i \frac{N_i}{V^{i-1}} + \frac{N}{V}\frac{dV}{dt}, \qquad (1)$$

where N is the number of particles in the volume V of the interaction space (which is, in general, time-dependent, i.e., variable), and k_i is the difference between the kinetic coefficients of production and loss of i particles.

If ECSs move over the surface randomly, the interaction space is represented by a plane (two-dimensional space), which has the dimension of area. When ECSs move regularly along the GCS boundary, the interaction space is a line (one-dimensional space). Its dimension is length and the interaction space volume equals the GCS perimeter (L).

In what follows, we will considering one- and two-particle processes in a GCS. The rate of change of the GCS perimeter (L) is given by

$$\frac{dL}{dt} = 2\pi v d\frac{N}{L} - \kappa = W(N, L), \qquad (2)$$

where d is the width of a metal zone arising as a result of scale removal for a single pass of ECSs along the contour line of the GCS, v is the average rate of ECS motion along the boundary, and κ is a parameter governing the rate of molten scale flow into the GCS.

It is seen that the system of Eqs. (1) and (2) does not have nontrivial general stationary solutions; i.e., it does not have a point of equilibrium and is unstable.

Thus, it seems that some essential factor has not been taken into account in the one-dimensional model of a GCS. Therefore, the model should be refined. Analysis of energy consumption and of the microrelief development in a GCS has led us to the assumption that two kinds of ECSs are present in a GCS.

Kind-1 ECSs, which are constantly located at the metal–scale interface, clean the surface. The other exists on the clean metal surface and does not participate in the cleaning process. Probably, ECSs of the sec-

1063-7842/02/4709-1202$22.00

ond kind are responsible for the microrelief inside a one-dimensional GCS in the form of a giant protrusion.

Note that the coexistence of the ECSs in these regions is possible only under specific conditions. We assume that they may arise when a near-surface plasma is generated on the clean metal surface by boundary ECSs. The plasma density must be sufficient to raise the probability of ECS production in this region and make it comparable with that at the metal–scale interface. Kind-2 ECSs form and exist within a cleaned metal belt of width h around the metal–scale interface. A certain critical density of the near-surface plasma over the cleaned metal surface (n^*) below which ECSs on the cleaned surface are not produced specifies the width h of the region where kind-2 ECSs exist.

Thus, the interaction between ECSs of both kinds leads to a nonlinear relationship between N and L, gives rise to stationary solutions to Eqs. (1) and (2), and, eventually, provides the existence of a stable GCS.

The relation between h and the phase variables is easy to find:

$$h = \frac{\pi}{L} C\chi \frac{N}{\dot{n}} + \frac{2\pi^3}{L^3}\left(C\chi\frac{N}{\dot{n}}\right)^2. \quad (3)$$

Here, χ is the ratio of the number of kind-1 ECSs to the total number of ECSs inside a GCS.

Next, we write the kinetic equation for kind-1 ECSs,

$$\chi\frac{dN}{dt} = k_1^l \chi N + k_2^l \chi^2 \frac{N^2}{L} + \chi\frac{N}{L}W(N, L), \quad (4)$$

and for kind-2,

$$(1-\chi)\frac{dN}{dt} = k_1^s(1-\chi)N + k_2^s(1-\chi)^2\frac{N^2}{Lh} + (1-\chi)\frac{N}{L}W(N, L). \quad (5)$$

Here, the subscripts at the coefficients k stand for the order of the reactions of ECS production and loss as in Eq. (1), and the superscripts l and s denote one-dimensional kinetics (one-dimensional space) and two-dimensional kinetics (two-dimensional space), respectively. The general kinetic equation for all the ECSs inside a GCS has the form

$$\frac{dN}{dt} = \alpha N + \beta\frac{N^2}{L} + \gamma\frac{N^2}{\xi + \frac{2\pi}{L^2}\xi^2 N} + \frac{N}{L}W(N, L) = F(N, L), \quad (6)$$

where

$$\alpha = \chi k_1^l + (1-\chi)k_1^s, \quad \beta = \chi^2 k_2^l,$$
$$\gamma = (1-\chi)^2 k_2^s, \quad \xi = C\pi\chi\frac{1}{\dot{n}}.$$

ESTIMATES OF THE KINETIC CONSTANTS OF ECSs

The practical value of the equations derived lies in the fact that, with certain experimental data available, they enable one to evaluate the kinetic constants of ECSs, i.e., in a sense, to solve the inverse problem.

To solve this problem, we used (in addition to the fact that a GCS can be stable) the following experimental data obtained earlier [2]: stationary values of the number of ECSs and sizes of GCSs; dependence of the GCS sizes on the number of ECSs in a single GCS; qualitative data for the microrelief within a stable GCS; estimate of the time of GCS relaxation to its stationary state; and quantitative data for the efficiency of surface cleaning by freely moving GCSs.

The results of our estimates are the following: $\alpha = 8.415 \times 10^3$ s^{-1}, $\gamma = 0.033$ cm^2/s, $(1-\chi)k_1^s = -8.4 \times 10^2$ s^{-1}, and $\chi k_1^l = 9.26 \times 1063$ s^{-1}.

To independently check these quantitative data, we took I–V characteristics of a single GCS in many measurements. A linear slowly increasing dependence with the slope

$$\frac{dU}{dI} = 0.15 \pm 0.02 \text{ V/A}$$

was observed. Data from [4] and the kinetic constants estimated make it possible to calculate this value independently. It was found that

$$\frac{dU}{dI} = 0.23 \text{ V/A}.$$

In view of the above assumptions, the agreement is quite reasonable.

CONCLUSIONS

(1) The possibility of the GCS stable state is explained by the existence of two interacting kinds of ECSs in a GCS. ECSs of the first kind are located at the metal–scale interface, which bounds the GCS. ECSs of the second exist on the cleaned metal surface within a belt around the interface. The width of the belt depends on the number of kind-1 ECSs.

(2) Estimates of the kinetic ECS coefficients were made.

(3) A special feature of kind-1 ECS kinetics was established. Among one-particle processes, the ECS production dominates over ECS loss, while for two-particle processes, the reverse situation takes place. As a whole, kind-1 ECSs make a major contribution to the kinetics of ECS formation in a GCS.

(4) A special feature of the kind-2 ECS kinetics is as follows. Among one-particle processes, ECS loss dominates over their production, while for two-particle processes, the situation is reverse.

REFERENCES

1. A. M. Mirkarimov, S. L. Pozharov, and I. V. Soldatov, Uzb. Fiz. Zh., No. 4, 45 (1996).
2. S. L. Pozharov, A. M. Mirkarimov, and I. V. Soldatov, Zh. Tekh. Fiz. **68** (11), 57 (1998) [Tech. Phys. **43**, 1323 (1998)].
3. U. A. Arifov *et al.*, European Patent No. 0468110 (1990); S. L. Pozharov *et al.*, RF Patent No. 93003651/12 (1993).
4. S. L. Pozharov, Uzb. Fiz. Zh. **2** (3), 247 (2000).

Translated by Yu. Vishnyakov