Theory of dynamic correspondence

G. E. Skvortsov

St. Petersburg State University, 199034 St. Petersburg, Russia
(Submitted October 9, 1998)
Zh. Tekh. Fiz. **69**, 1–6 (October 1999)

A general theory of dynamic similarity, which includes processes with strong disequilibrium and considers the generalized similarity of processes, is discussed. The basic laws of the theory, viz., the measures of action and dynamic correspondence, are formulated. A basic tool for dynamic correspondence, viz., the object–process classifier, is proposed. An initial description of the principal parts of the theory is given. Possible effects and the relations describing them are indicated. © *1999 American Institute of Physics.* [S1063-7842(99)00110-5]

The theory of dynamic similarity is an irreplaceable first step in the analysis of processes. The existing theory[1] covers the region of moderate disequilibrium, but does not extend directly to processes far from equilibrium, since, generally speaking, there is no similarity between processes in that region. A generalized theory is needed wherein the fundamental relations would be given in the form of quite general and simple semiphenomenological expressions that could satisfactorily describe the processes in diverse systems.

This paper is devoted to the important problem of devising a theory of dynamic correspondence covering the region of strong disequilibrium. Dynamic correspondence signifies the establishment of an analogy between processes with respect to specific characteristics: an analogy between dynamical equations, regimes, mechanisms, and the analytical forms of fundamental relations. Although regularities (conformities) can serve as the most general indicators, one needs analytical expressions in order for them to be constructive.

In this paper we combine fragmentary results: the laws, principles, and methods which we derived in Refs. 2–7. Laws governing measures of action and dynamic correspondence are formulated, and their consequences and applications are given. A novel tool, viz., the object–process classifier, is employed to determine dynamic correspondence and to predict new effects and regularities. The general form of fundamental regularities is presented as a part of a theory of dynamic correspondence for regimes with different degrees of disequilibrium (regular, resonant, and structural transitions).

1. INITIAL POSTULATES

Processes with a broad range of disequilibrium caused by various disturbances in diverse systems are considered. The system changes its structure during these processes, so that we must treat a system/process (a dynamic system).

a) The structure of a system is formed by structural kinetic elements (SKEs) and bonds, i.e., the interactions between them.

As we know, a structure has a hierarchical composition: a first-level SKE consists of second-level SKEs, etc. A structure generally has stochastic properties, and the mean characteristics of the SKEs are then used. We assume that a structure has s_m levels and is described by a set of parameters

$$\{n_s, \mu_s, \lambda_s, \tau_s, v_s, \varepsilon_s, q_s\}, \quad s=1,2,\ldots,s_m, \tag{1}$$

where n_s is the density, μ_s is the mass, λ_s and τ_s are the characteristic dimension and time of the SKEs, v_s is the root-mean-square velocity, ε_s is the interaction (bond) energy, and q_s is the charge, as well as additional characteristics of the SKEs of the level s; theses characteristics depend on the external conditions and intensity of the disturbances.

In the case under consideration, a structure which matches the scales and energetics of the process is chosen from the complete existing hierarchy as the fundamental structure. Depending on the ranges of their values in physical systems, the original SKEs can be elementary particles, atoms, molecules, quasiparticles, fluctons of various kinds, and various macroscopic formations.

The structure determines the basic properties of a dynamic system and is reflected in the dynamic equations. In the case of the classical, weakly nonequilibrium theory, the structure is represented by kinetic coefficients. For a strongly nonequilibrium theory, or superhydrodynamics,[7–9] as a consequence of the structural conditionality the equations contain nonlinear functional relations with delay and nonlocality.

b) To study and compare processes in different systems, we must have a set of suitable, fairly universal action measures, which define the dynamic state of the system. Such a set is obtained on the basis of the set of macroscopic quantities appearing in the general dynamic equations $\{a_n(x,t)\}$ (Refs. 7 and 8) using reliable structural characteristics of the dynamic system. In this case suitable complexes are formed with consideration of the natural conditions: positivity, a tendency to rise as the action factor increases, and ''normalization to the transition.''

The action measures are the ratios between the action factors $g(x,t)$ and the corresponding characteristics of the system, i.e., the structure factors $s[g]$,

$$G=\frac{g}{s[g]}: \quad A_n=\left|\frac{a_n-a_{n0}}{a_{ns}}\right|:$$

$$C=\left|\frac{\rho_n-\rho_{n0}}{\rho_{ns}}\right|, \quad V=\left|\frac{u_n-u_{n0}}{u_{ns}}\right|,$$

$$E=\left|\frac{e_n-e_{n0}}{\varepsilon_s}\right|, \quad N=\left|\frac{\dot{e}_n-\dot{e}_{n0}}{\dot{\varepsilon}_s}\right|,$$

$$J=\left|\frac{j_n-j_{n0}}{j_{ns}}\right|, \quad F=\left|\frac{f_n}{f_s}\right|,\ldots, \tag{2.1}$$

$$H=\tau_s|\partial_t \ln|a_n||, \quad L=\lambda_s|\partial_x \ln|a_n||,\ldots, \tag{2.2}$$

where $a_{n0}(x,t)$ are the defining parameters of the comparison process; $a_{ns}(x,t)$ are the corresponding structural characteristics; ρ_n are the densities; u_n are the velocities; e_n, $\dot{e}_n$, and f_n are the energies, powers, and forces of the disturbance acting on the SKEs; j_n are the fluxes of a_n; $\dot{\varepsilon}_s/\tau_s$; $j_{ns}\sim v_{ns}a_{ns}$; and $f_s\sim\mu_s v_s/\tau_s$.

The special disturbance regimes, i.e., resonances and structural transitions, are taken into account by employing action measures of the form

$$R_G=\sum_m \frac{b_m}{\left[1+k_m^{\pm}\left(\frac{G-G_{mc}}{G_{mc}}\right)^{\gamma_m^{\pm}}\right]}, \tag{3}$$

where G_{mc} are the threshold values of G_m and the superscripts $\pm$ indicate values to the right and the left of the threshold (for resonances without a qualitative change in the structure $k_m^-=k_m^+$ and $\gamma_m^{\pm}=2$).

The quantities defined by (2) and (3) should be used together with the action and structure factors calculated relative to the original structure

$$G_0(x,t)=\frac{g}{s_0}, \quad S=\frac{s[g]-s_0}{s_0}, \quad s_0=s[0]. \tag{4}$$

The quantities defined by (1)–(4) make up the complete set of action measures.

The active action measures (2.1), to which the arrival or escape of resource quantities (mass, momentum, and energy) are related, and the scaling action measures (2.2), which specify the relations between internal and external temporal and spatial scales, can be separated according to the nature of the action factors.

It should be pointed out that the use of an action measure permits consideration of a more extensive list of reference systems and processes than within the degree of disequilibrium, which measures the deviation from quasiequilibrium. This circumstance is of great significance for processes in chemistry and biology, as well as for disturbances acting on structural transitions. For example, according to a basic law of living systems,[10] external disturbances diminish the disequilibrium inherent to them (the action measure increases, and the disequilibrium decreases). Determination of the disequilibrium for complete transfer through a membrane is instructive.[11]

The meaning of this part of the theory can be stated concisely as the action measure law: ''the result of a disturbance acting on a system reflects a set of universal action measures; each action measure reveals a mutual conditionality between structure and action.''

A variant of the action measure law for biological systems is the initial level law.[12]

c) The complete, universal set of action measures and the system of laws in Refs. 2–7 serves as the fundamental law of dynamic correspondence: ''processes in diverse system have something in common, whose composition and extent are determined by the system of action measures.''

With consideration of the dependence of action measures on structure, this law can be regarded as a dynamic generalization of the familiar equilibrium law of the similarity of structure and properties (Mendeleev's periodic law, Vavilov's law of homologous series, etc.).

The dynamic correspondence principle, which points out characteristics and methods for comparing processes, should be used to constructively compare processes. The basic characteristics are: the similarities of equations, regimes, functional relations of the defining parameters, mechanisms, and effects. The basic method for revealing dynamic correspondence is object–process classification.

d) This method is implemented by an object–process classifier, which is an ordered system of concentrated information on classes of objects, possible processes caused by disturbances of various kinds, and their regularities. For objects in physics with one disturbance and three action measures (the amplitude A and the scales H and L) the object–process classifier has the form of a table with an axis for object classes (..., molecules, fields, gases, etc.) and two axes for the action measures, viz., A and $H\simeq L$.

The dynamic density of the object–process classifier contains fundamental information from the following areas: the equilibrium structure and properties of each class of objects ($G=0$), thermodynamics ($A<1, H=0$), classical hydrothermoelectrodynamics ($A<1, H<0.2$), synergetics ($A\leqslant 1, H<0.2$), superhydrodynamics ($A\leqslant 1, H<10$), resonance regimes ($A<0.2, H=1$), structural transitions ($A=1$), anomalous regimes ($A_0\gtrsim 1$), hydrodynamics-2, synergetics-2, etc. The areas with the number 2 contain dynamic systems obtained from the systems indicated as a result of an above-threshold disturbance; for example, ''hydrodynamics-2'' refers to the theory of turbulized or boiling liquids.

The concentrated information on the processes includes the fundamental effects controlling the laws and regularities, as well as the form of the equations.

An object–process classifier, being a kind of dynamical analog of the periodic table, has some significant properties. We first note the universal classification property, which establishes the correspondence between classes of processes and orders the regimes of processes. An instructive example is provided by an analysis of the electrospraying of liquids.[13] The prognostic property, which permits the prediction of new effects and regularities, is important. For example, it is possible to predict the variety of effects of a structural transition (in particular, of structural transitions of fields), anomalous effects, and ''windows of quasiequilibrium'' (Ref. 14).

It can be stated on the basis of the prognostic property that the phenomenon of anomalous relaxation discovered in gases[15] also occurs in liquids and solids.

Using the principle of dynamic correspondence and object–process classification, we can characterize the synergetics and superhydrodynamics of a multinucleon nucleus, a macromolecule, and a field.

2. PROPERTIES AND EFFECTS OF STRUCTURAL CONDITIONALITY

Let us consider some significant properties of dynamic correspondence and effects which are determined by the structure.

a) The form of the action measures (2) leads to several useful conclusions. An equality between the action measures for processes with the same action factor in two systems (or modifications of a single system) signifies their complete correspondence, i.e., similarity:

$$G_1 = G_2, \tag{5}$$

$$\frac{g_1}{g_2} = \frac{s_1[g_1]}{s_2[g_2]} \sim \frac{s_{10}+s_{11}g_1}{s_{20}+s_{21}g_2}. \tag{6}$$

The latter correspondence goes over to an equality when $g_1 \sim g_2 \to 0$. It follows from (5) and (6) that a small disturbance can produce an identical result in a system having a proportionally smaller structure factor. Smaller values of the structure factors (velocities of sound, binding energies, reciprocal relaxation times, and structural dimensions) correspond to dynamic states in which the system has a large number of degrees of freedom.

According to (2), for equal values of g, dynamic systems with a larger number of degrees of freedom are farther from equilibrium. This conclusion accounts for the increased disequilibrium of a gas of excited particles, a turbulent liquid, and polymer systems.

In the limit of small G it is expedient to compare different systems using the ''potential disequilibrium'' measure $s_{12} = s_{20}/s_{10}$ (5), which has an obvious relationship to the number of degrees of freedom and complexity of the system. Essentially this measure was used to classify liquid media in Ref. 8.

b) The condition that the action measures be ''normalized to the transition'' means that the equality of G to unity reflects the law of the quality boundary and defines it:

$$g_c \approx s[g_c]: \quad \Delta\rho \approx \rho_c, \quad \Delta u \approx v_s,$$
$$\Delta e \approx \varepsilon_s, \quad \Delta j \approx j_s, \ldots; \tag{7}$$

$$|\partial_t \ln \Delta a| \approx 1/\tau_s, \quad |\partial_x \ln \Delta a| \approx 1/\lambda_s, \ldots . \tag{8}$$

The first condition in (7) defines a concentration quality boundary—a phase transition, an ignition boundary. The second condition, involving the sound velocity v_s, corresponds, as we know, to the appearance of shock waves. The third condition signifies disintegration of the SKEs of level s: fragmentation, dissociation, and anomalous relaxation.[2] The fourth condition indicates the critical flux, the attainment of which is accompanied by the formation of a new structure, particularly, a dissipative structure, which can be treated by synergetics.

The first condition in (8) corresponds to frequency resonance; with consideration of Le Chatelier's principle, it also gives the simple relaxation equation $d_t \Delta a = -\Delta a/\tau_s$. The second condition indicates the wavelength with the greatest relative absorption. These conditions can be regarded as estimates of parameters of the structures formed as functions of the rate or gradient of the disturbance. For example, when a melt is cooled at the rate

$$\left| \partial_t \ln \left| \frac{T - T_c}{T_c} \right| \right|,$$

where T_c is the solidification point, the first condition gives an estimate of λ_s for the SKEs formed:

$$\lambda_s \simeq v_s / |\partial_t \ln \Delta T|, \tag{9}$$

where v_s is the velocity of the solidification boundary.

The gradient analogs of the velocity and power action measures were not listed among the action measures (2); they are obtained by multiplying the action measure (2.1) by the gradient action measure (2.2). For example, the condition for the appearance of turbulence has the form[16]

$$V_y / L_y = \frac{y_*^2}{\nu} \frac{dv_x(y)}{dy} \approx 1, \tag{10}$$

where y_* is the distance from the wall in the transverse direction to the flow toward the region where turbons are generated.

c) The structural characteristics determining the action measures are adequately represented in the expressions for nonequilibrium kinetic coefficients. When included in the theory of dynamic correspondence, the general expressions for the nonequilibrium kinetic coefficients[8,9] make it possible to reveal the correspondence of nonequilibrium kinetic coefficients, to determine structural characteristics, and to point out numerous structural kinetic effects.

We use the general expressions for the nonequilibrium viscosity and conductivity[9] and obtain the corresponding approximate relations

$$\eta \sim \frac{\mu_e v_T}{\sigma_e (1-\gamma)\left[Z + 1 + \dfrac{\tau_2}{\tau_1} + F \right]},$$

$$\sigma \sim \frac{n_q q^2}{n \mu_q v_T \sigma_e[\ldots]}, \tag{11}$$

where

$$\tau_1 = (n_e v_T \sigma_e)^{-1}, \quad \gamma \simeq \pi \lambda_2^3 n_e,$$
$$Z \equiv z + ikV, \quad F = \tau_1 f / \mu_1 v_T,$$

and $\sigma_e(T,F) \sim T^{-r}$ is the collision cross section of the SKEs.

As is clear from (11), the features of the behavior of these coefficients are stipulated by the features of the structural characteristics. These features can be significantly

smaller or greater values (in comparison to the mean value for a given class of systems), as well as abrupt changes (jumps) in the characteristics.

For example, the tendency of the collision frequency to approach zero at high energies leads to a decrease in the relaxation rate in proportion to $\exp(-t/\tau_s)^{\alpha}$, where $\alpha<1$ (Ref. 17). Such a ''protracted'' exponential function describes relaxation in disordered systems (glasses).[18]

In the case of shock waves there are data indicating an abrupt increase in the relaxation time (during an overall decrease) as the shock-wave velocity rises.[19] According to this feature, the corresponding regime for $u\approx 9$ km/s in air should be considered anomalous. It is close to the instability regime of shock waves indicated in Ref. 20.

Abrupt variation of the mass of the SKEs causes various nonequilibrium effects. For example, rapid dissociation of a gas leads to a decrease in viscosity according to (11): $\eta \sim \mu^{1/3}$. The Gunn effect is known to be caused by an abrupt increase in the effective mass $\mu(E)$ at $E>E_c$.

3. GENERAL SIMILARITY OF NONEQUILIBRIUM RELATIONS

The principle of dynamic correspondence and the general expressions of the defining relations[8,9] can be used to discuss the similarity of the form of fundamental nonequilibrium relations over a broad range of variation of the action measures. We shall consider fundamental relations which reflect the reaction of a system to disturbances: the response of the system and the relaxation, transport, and resistance relations (laws):

$$R_{\alpha}=\Phi_{\alpha}[G], \quad G=(A,H,L,F), \tag{12}$$

where $\Phi_{\alpha}[.]$ are nonlinear operators of the functional type, whose general expressions were given in Ref. 8.

a) For small action measures ($G<0.2$) relations (12) take the form of the linear relations

$$\Delta R_{\alpha}=\Phi_{\alpha}[G]-\Phi_{\alpha}[0]\simeq\sum_{\beta} K_{\alpha\beta}G_{0\beta}. \tag{13}$$

These relations are suitable for any systems and describe the linear response, Ohm's law, Fourier's law, Stokes' law, etc. The susceptibility coefficients in (13) are determined by the structural characteristics of the systems at equilibrium.

In the case of small disturbances ($G_0\ll 1$) acting on strongly nonequilibrium states, for example, on a structural transition, relations (12) take the (scaling) form

$$\Delta R_{\alpha}\simeq\sum_{\beta} K_{\alpha\beta}G_{0\beta}^{r_{\beta}}, \quad r_{\beta}<1. \tag{14}$$

b) for regimes with $G_A<0.2$ and $H, L>0.3$ the relations (12) reduce to linear functionals, i.e., convolutions in time and space, which take the following form in the Laplace–Fourier representation

$$\Delta R_{\alpha}\simeq\sum K_{\alpha\beta}(z,k,\Pi)G_{\beta}(z,k), \tag{15}$$

where $z\to\tau_s\partial_t=H$ and $ik\to\lambda_s\nabla=L$.

Expressions for the nonequilibrium susceptibilities $K_{\alpha\beta}(z,k)$ have been obtained in a form suitable for constructing a semiphenomenological theory for a gas and any z and k (Refs. 8 and 9). We note that the derivation of the asymptotic relations for $|z,k|\gg 1$, which are needed for constructing effective broad-range interpolation expressions, was the main result of that work.

c) We point out the asymptotic nature of nonequilibrium kinetic coefficients. As follows from an analysis of a series of problems, we have

$$K(z,ik)\simeq\frac{c_1+ic_2}{c_3z^{1+\alpha}+|k|^{1+\gamma}}, \tag{16}$$

where $|\alpha,\gamma|<0.5$ (as a rule).

In the case of a linear theory without structural transitions, $\alpha=\beta=0$ (Refs. 8 and 9). For type-II superconductors $\gamma=0$ in the stationary case (failure to take the modulus of k in Ref. 21 leads to an incorrect form of the field penetration constant).

d) Using the correspondence $k\to L$, we can apply the result (16) to the nonlinear case. In the problem of simple shear[22] the shear viscosity has the form

$$\eta=c\left|\frac{\lambda_1}{\nu_1}\partial_y u_x\right|^{-4/3}, \quad \text{i.e.} \quad \gamma=\frac{1}{3}.$$

In the theory of turbulence constructed on a rational basis in Ref. 23, where $\mathrm{Re}\sim 10^6$, the turbulent viscosity decreases with increasing values of the gradient action measure (whose form was obtained) with an exponent equal to 2/3, i.e., $\gamma=-(1/3)$. The familiar Kármán phenomenological theory gives $\gamma=0$ for $\mathrm{Re}>10^8$.

e) Fairly universal expressions for action measures of the active type can be obtained only with confinement to the stability limit of the structure. In particular, for the conductivity in a constant electric field, the following relation is obtained from the general expression when $\sigma_e \sim(v_T/v_0)^{-2s}$ and $n_e=\mathrm{const}$ (Ref. 24):

$$\sigma(F)\simeq cF^{-(1+\gamma)}. \tag{17}$$

If the structural-transition boundaries are separated considerably ($G_{01}\ll G_{02}\ll G_{03},\ldots$), on a general basis we obtain

$$K(G)\sim c_nG_0^{-(1+\gamma_n)}, \tag{18}$$

where $|\gamma_n\leqslant 1|$.

The converse of this result is also true, i.e., a conclusion that a structural transition exists can be drawn on the basis of a significant change in the power over a small range of values of G_0.

f) It would be of interest to describe regimes corresponding to $G_{A0}<1.3$, $G_A\lesssim 1$, and $H, L<0.2$. Fairly universal relations were given for this region in Ref. 5, and a series of concrete problems was treated on their basis.

Within the theory of dynamic correspondence it is expedient to use the following more general and fairly simple interpolation expression, which has been reconciled with all the laws of disequilibrium, the reactions of systems, and the regularities of structural transitions:[2–6]

$$\Phi(x)=\frac{b_3x^3+b_2x^2+b_1x+b_0}{c_3x^3+c_2x^2+c_1x+c_0}\equiv\frac{B(x)}{C(x)}, \quad x=G_0. \tag{19}$$

This single-factor expression with seven parameters describes the principal regimes: linear variation, nonlinear (monotonic) variation, resonance, and second- and first-order structural transitions. It is suitable for diverse systems (from atomic to cosmic) and has been tested on tens of problems.

Relation (19) can be simplified for each of the regimes by setting the appropriate coefficients equal to zero and assigning specialized forms to others. For example, for resonance we have $b_3=b_2=b_1=c_3=0$, $c_1=-2x_cc_2$, and $c_0=c_2x_c^2+\varepsilon^2$, where x_c and ε are parameters of the Lorentzian.

The features of the behavior of the functions reflect zeros of the derivative, which are determined for (19) by the numerator:

$$\Phi'(x)=\frac{\Sigma_{n=0}^4 d_n x^n}{C^2(x)}\equiv\frac{D(x)}{C^2(x)},$$

$$d_4=b_3c_2-b_2c_3,\quad d_3=2(b_3c_1-b_1c_3),$$

$$d_2=3(b_3c_0-b_0c_3)+b_2c_1-b_1c_2,$$

$$d_1=2(b_2c_0-b_0c_2),\quad d_0=b_1c_0-b_0c_1. \tag{20}$$

In the general case $\Phi(x)$ is nonmonotonic, and the equation $D(x)=0$ has two or four real roots. If there are four positive roots, $D(x)$ is represented in the form

$$D(x)=d_4\prod_{n=1}^{4}(x-x_n),\quad x_n>0.$$

In the case where the roots are fairly separated, they are approximately equal to

$$x_n\approx-\frac{d_{n-1}}{d_n}, \tag{21}$$

where x_n are the boundaries of two consecutive first-order structural transitions.

It follows from the positivity of $\Phi(x)$ that

$$d_{0,2,4}>0,\quad d_{1,3}<0\rightarrow b_3c_2>b_2c_3,\quad b_3c_1<b_1c_3. \tag{22}$$

Such constraints simplify selection of the parameters.

If two roots coincide, we have a second-order structural transition or a critical point (and the corresponding constraints on the parameters). In the case of $c_3=b_0=0$, for two linear regimes with $x\ll 1$ and $x>x_m$, where x_m is the maximum root, under the condition $b_1/c_0>b_3/c_2$ the susceptibility of the system increases as a result of the structural transition (or decreases under the reverse condition), and it can be stated that we have an activating (or deactivating) structural transition. The foregoing partial analysis demonstrates the favorable possibilities of using (19).

g) The overall functional form of the defining relations obtained on the basis of operator expressions[8] in accordance with the principle of dynamic correspondence has the form

$$J_n[G]=\sum_{l,s,m}\int_0^t dt'\int dx'\,\varphi_{nm}^{ls}(\Delta t,\Delta x;G,G')\times\exp\left[-H_{nm}^s\left(\frac{|\Delta t|}{\tau_{nm}^s(G')}\right)-L_{nm}^l\left(\frac{|\Delta x|}{\chi_{nm}^l(G')}\right)\right]G_m', \tag{23}$$

where $\Delta t=t-t'$, $\Delta x=x-x'$, $G'\equiv G(t',x')$, and φ, H, and L are functions, whose behavior can be indicated for small and large values of the arguments.

The principle of dynamic correspondence in (23) means that the specialized forms of this expression should correspond to all existing theoretical and experimental data. Such correspondence can be demonstrated for all the regimes considered above. However, a large amount of experimental data, particularly on the dynamics (rheology) of polymers, must be compared and described using the functional relation (23).

4. DISCUSSION OF RESULTS

The results of the present work and our preceding studies[2–8] permits expansion of the theory of dynamic similarity presented in the notable monograph by S. S. Kutateladze[1] and inclusion of processes far from equilibrium in the analysis. We shall briefly discuss the principal theories of dynamic correspondence, comparing the results of this work and the material in Ref. 1. Such a comparison is very instructive and fruitful. As we know, a theory is made up of four parts: regularities, models, methods, and demonstration problems. The main problem in the theory of dynamic correspondence is to find some common ground in the diversity of processes and to reflect that common ground by means of functional relations.

1) The conventional foundation of the theory of dynamic correspondence consists of the postulate of the unity of nature, a system of laws and regularities,[2–7] other known laws, and the principle of invariance of the form of the fundamental relations with respect to the system of measure. The traditional foundation of dynamic similarity in the form of the familiar equations of hydrothermoelectrodynamics, which describe moderately nonequilibrium processes, was used in Ref. 1.

2) The models of the theory of dynamic correspondence cover very general mechanisms of processes, as well as analytical models for describing processes. The general mechanisms of processes with moderate disequilibrium are provided by synergetics. Analytical models are obtained from a dimensionality analysis, asymptotic relations of the processes, and fairly general expressions like those in Sec. 3, which have been reconciled with the system of laws. The forms of analytical models of various kinds were given in Ref. 1. In particular, the relation for the thermal conductivity of metals given by (13) and (17) is consistent with (19) when $b_3=b_2=b_0=0$, $b_1=3$, $c_2=c_1=0$, $c_3=2$, and $c_0=1$. We note that the occurrence of a structural transition at $\overline{\lambda}_{\max}$, $\overline{T}\approx 1$ can be concluded from the correspondence between a structural transition and an extremum on a defining relation.

3) The methods of the theory of dynamic correspondence were partially described in Ref. 1. The principal tool for defining the general nature of processes is the object–process classifier presented in Sec. 1 d. A suitable method for analytical description is to obtain an increasingly more general form of the fundamental relations, which is reconciled with the laws. Examples of such an approach are provided by Eqs. (19) and (23).

4) A comparison of the general model of structure used in Refs. 2–7 and the arguments for choosing a structure, as well as the examples of such choices in Ref. 1, attests to the closeness of understanding this question and the expediency of applying a general model to the theory of dynamic correspondence.

5) Numerous, very interesting demonstration models of the theory of dynamic correspondence were given in Ref. 1. They have been supplemented by the problems considered in Refs. 2–9. The fruitfulness of the two approaches becomes clearly evident in the example of the problem of a boiling liquid[1] considered from the standpoint of describing a structural transition.[5]

[1] S. S. Kutateladze, *Analysis of Similarity and Physical Models* [in Russian], Novosibirsk (1986), 263 pp.
[2] G. E. Skvortsov, Pis'ma Zh. Tekh. Fiz. **16**(17), 15 (1990) [Sov. Tech. Phys. Lett. **16**(9), 647 (1990)].
[3] G. E. Skvortsov, Pis'ma Zh. Tekh. Fiz. **23**(6), 85 (1997) [Tech. Phys. Lett. **23**(3), 246 (1997)].
[4] G. E. Skvortsov, Pis'ma Zh. Tekh. Fiz. **23**(7), 23 (1997) [Tech. Phys. Lett. **23**(4), 261 (1997)]; Pis'ma Zh. Tekh. Fiz. **23**(10), 17 (1997) [Tech. Phys. Lett. **23**(5), 383 (1997)].
[5] G. E. Skvortsov, Pis'ma Zh. Tekh. Fiz. **24**(3), 80 (1998) [Tech. Phys. Lett. **24**(2), 117 (1998)]; Pis'ma Zh. Tekh. Fiz. **24**(19), 7 (1998) [Tech. Phys. Lett. **24**(19), 749 (1998)].
[6] G. E. Skvortsov, Pis'ma Zh. Tekh. Fiz. **25**(1), 81 (1999) [Tech. Phys. Lett. **25**(1), 35 (1999)].
[7] G. E. Skvortsov, Pis'ma Zh. Tekh. Fiz. **23**(22), 7 (1997) [Tech. Phys. Lett. **23**(11), 861 (1997)].
[8] G. E. Skvortsov, Vestn. Leningr. Univ., No. 3, pp. 94–98 (1979).
[9] G. E. Skvortsov, Zh. Éksp. Teor. Fiz. **63**, 502 (1972) [Sov. Phys. JETP **36**, 266 (1973)]; Zh. Éksp. Teor. Fiz. **68**, 956 (1975) [Sov. Phys. JETP **41**, 473 (1975)].
[10] É. S. Bauér, *Theoretical Biology* [in Russian], Budapest (1982), 295 pp.
[11] V. F. Pastushenko, Kh. Shindler, and Yu. A. Chizmadzhev, Dokl. Akad. Nauk SSSR **320**(6), 1491 (1991).
[12] N. V. Vasil'ev and T. I. Kolyada, *Living Systems under an External Disturbance* [in Russian], St. Petersburg (1992), 440 pp.
[13] S. O. Shiryaeva, A. I. Grigor'ev, and L. S. Podval'nyĭ, Pis'ma Zh. Tekh. Fiz. **19**(4), 36 (1993) [Tech. Phys. Lett. **19**(2), 116 (1993)].
[14] A. Yu. Ivanov, Pis'ma Zh. Tekh. Fiz. **25**(8), 1 (1999) [Tech. Phys. Lett. **25**(4), 293 (1999)].
[15] G. I. Mishin, A. P. Bedin, and N. I. Yushchenkova *et al.*, Zh. Tekh. Fiz. **51**, 2315 (1981) [Sov. Phys. Tech. Phys. **26**, 1363 (1981)].
[16] B. S. Petukhov, N. F. Polyakov, and Yu. V. Tsipulyaev, Teplofiz. Vys. Temp. **16**, 987 (1978) [High Temp. **16**, 839 (1978)].
[17] G. E. Skvortsov, Zh. Éksp. Teor. Fiz. **57**, 2054 (1969) [Sov. Phys. JETP **30**, 1114 (1970)].
[18] T. Riste and D. Sherrington (Eds.), *Phase Transitions and Relaxation in Systems with Competing Energy Scales*, Kluwer, Dordrecht–Boston (1993), 452 pp.
[19] M. B. Zheleznyak and A. Kh. Mnatsakyanyan, Teplofiz. Vys. Temp. **6**, 390 (1968).
[20] A. S. Baryshnikov and G. E. Skvortsov, Zh. Tekh. Fiz. **49**, 2483 (1979) [Sov. Phys. Tech. Phys. **24**, 1401 (1979)].
[21] E. A. Turov, *Material Equations of Electrodynamics* [in Russian], Moscow (1983), 156 pp.
[22] A. Santos, V. Garzo, J. Brey, and J. W. Dufty, Phys. Rev. Lett. **71**, 3971 (1993).
[23] G. E. Skvortsov, Zh. Tekh. Fiz. **59**(3), 62 (1989) [Sov. Phys. Tech. Phys. **34**, 289 (1989)].
[24] E. N. Perevoznikov and G. E. Skvortsov, Zh. Tekh. Fiz. **61**(9), 1 (1991) [Sov. Phys. Tech. Phys. **36**, 967 (1991)].

Translated by P. Shelnitz

Features of the bremsstrahlung accompanying collisions with a hydrogen atom in an excited state

A. V. Korol', O. I. Obolenskiĭ, and A. V. Solov'ev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted February 12, 1998; resubmitted March 16, 1999)
Zh. Tekh. Fiz. **69**, 7–13 (October 1999)

The features of the bremsstrahlung appearing during a collision of a fast charged particle with a hydrogen atom (or hydrogenic ion) in an excited state are investigated. It is shown that the emission spectrum of photons with energies greater than the ionization potential of a given excited state (except the $2s$ state) displays narrow lines, which are caused by de-excitation of the atom in an intermediate state. It is demonstrated that the scattering of a charged particle on an excited hydrogen atom produces a feature which is not observed in the case of scattering on a ground-state hydrogen atom. Expressions are obtained for the generalized dynamic polarizability of the hydrogen atom and hydrogenic ions in the $1s$, $2s$, and $3s$ states. A method is developed for deriving expressions for the generalized dynamic polarizabilities of other excited states through the use of the Coulomb Green's function and representation of the electronic wave function in terms of the differentiation of the generating functions of Laguerre polynomials. The bremsstrahlung cross sections for electrons and positrons colliding with hydrogen atoms in the $1s$, $2s$, and $3s$ states are calculated. [S1063-7842(99)00210-X]

INTRODUCTION

In this paper the bremsstrahlung cross sections for fast charged particles, i.e., electrons and positrons, colliding with a hydrogen atom in the ground state and in an excited state are calculated and compared with one another. The dependences of the cross sections of the process on the frequency of the photon emitted and the projectile velocity are investigated. Allowance is made for two principal mechanisms of photon emission, viz., ordinary[1,2] and polarization bremsstrahlung, which appears as a result of dynamic polarization of the target atom by the electric field of the projectile.[3–5] The important role of the interference of the two bremsstrahlung mechanisms is demonstrated.

The polarization bremsstrahlung accompanying the collisions of electrons, positrons, and protons with hydrogen atoms in the ground state was considered in Ref. 6–8. This paper examines the role of the polarization bremsstrahlung mechanism during collisions of charged particles with hydrogen atoms in excited states in detail. It is shown here that there are significant differences in the behavior of the bremsstrahlung following the transition from the ground state of the target to excited states.

The contribution of the polarization bremsstrahlung, as well as the interference term in the cross section, are calculated using the Coulomb Green's function. The wave functions of the initial and final states are obtained by differentiation with respect to a parameter of a simple generating function.[9] This procedure yields closed expressions for the dynamic response of a target atom for an arbitrary principal quantum number. The atomic system of units ($|e|=\hbar=m_e=1$) is employed in this work.

BREMSSTRAHLUNG CROSS SECTION DURING THE COLLISION OF A CHARGED PARTICLE WITH A HYDROGEN ATOM OR A HYDROGENIC ION

Let us consider a collision of a charged particle having a momentum $\mathbf{p}$ and an energy ε_1 with a hydrogen atom in the ground or excited state $|j\rangle$. The collision results in the emission of a photon with a frequency ω, and the energy and momentum of the projectile become equal to ε_2 and $\mathbf{p}'$, while the state of the target atom remains unchanged. The differential bremsstrahlung cross section for s states in the Born dipole approximation with consideration of the contributions of the ordinary and polarization mechanisms is described by the formula[3]

$$\frac{d^4\sigma}{d\omega dq d\Omega_k}=\frac{4}{3}\frac{\omega^3}{\pi c^3 v^2 q}\left|\frac{Z-W_j(q)}{\mu\omega^2}+e\alpha_j(\omega,q)\right|^2 \times\{1-P_2(\cos\theta_q)P_2(\cos\theta_k)\}. \quad (1)$$

Here ω is the frequency of the photon emitted, v is the projectile velocity, $c\approx 137$ is the speed of light, $q=|\mathbf{p}-\mathbf{p}'|$ is the momentum transferred during the collision, Z is the atomic number the target atom, e is the projectile charge (for an electron $e=-1$), μ is the reduced mass of the projectile and the target atom, $\theta_k=\widehat{\mathbf{kp}}$ is the angle between the photon escape direction and the direction of motion of the projectile, $\cos\theta_q=\cos\widehat{\mathbf{qp}}=(2\omega+q^2)/2vq$, $P_2(x)$ is a Legendre polynomial, $W_j(q)$ is the form factor of the state of the atom with the set of quantum numbers j, $W_j(q)=\int dV e^{i\mathbf{q}\cdot\mathbf{r}}|\Phi_j(\mathbf{r})|^2$, and $\alpha_j(\omega,q)$ is the generalized dynamic dipole polarizability of the atom:

$$\alpha_j(\omega,q)=-\frac{i}{q}\sum_f\left[\frac{\langle j|e^{-i\mathbf{q}\cdot\mathbf{r}}|f\rangle\langle f|\mathbf{r}\cdot\mathbf{n_q}|j\rangle}{\omega-\omega_{fj}+i0}-\frac{\langle f|\mathbf{r}\cdot\mathbf{n_q}|f\rangle\langle j|e^{-i\mathbf{q}\cdot\mathbf{r}}|j\rangle}{\omega+\omega_{fj}+i0}\right], \quad (2)$$

where $\omega_{fj}=E_f-E_j$ are the frequencies of the transitions between states with the energies ε_f and ε_j.

Integrating (1), we obtain the spectral dependence of the cross section

$$\frac{d\sigma}{d\omega}=\frac{16}{3}\frac{\omega^3}{c^3v^2}\int_{q_{\min}}^{q_{\max}}\frac{dq}{d}\left|\frac{Z-W_j(q)}{\mu\omega^2}+e\alpha_j(\omega,q)\right|^2, \quad (3)$$

where

$$q_{\min}=\mu v\left(1-\sqrt{1-\frac{2\omega}{\mu v^2}}\right), \quad q_{\max}=\mu v\left(1+\sqrt{1-\frac{2\omega}{\mu v^2}}\right). \quad (4)$$

In the cross sections (1) and (3) the term proportional to $(Z-W_j(q))/\mu\omega^2$ describes ordinary bremsstrahlung, and the term proportional to $e\alpha_j(\omega,q)$ describes polarization bremsstrahlung. Since the numerator in the ordinary bremsstrahlung amplitude contains the reduced mass μ, the ordinary bremsstrahlung mechanism is suppressed in the case of a heavy projectile, such as a proton. In the case of a collision of an electron or a positron, the two bremsstrahlung mechanisms make contributions of the same order of magnitude. In this case the recoil of the electron (positron) following the collision has a significant influence on the limits of the range of momentum transfers $q_{\min}$ and $q_{\max}$ at $\omega\sim v^2/2$, while the recoil can be neglected for a heavy particle in this frequency range.

CALCULATION OF THE GENERALIZED DYNAMIC DIPOLE POLARIZABILITY OF THE HYDROGEN ATOM IN AN EXCITED STATE USING THE COULOMB GREEN'S FUNCTION

The formulas presented above show that the behavior of the total bremsstrahlung cross section is determined by the generalized dynamic polarizability. For many-electron atoms obtaining the generalized polarizability is a complex problem requiring a numerical calculation.[5,10] An analytical solution of this problem is possible for hydrogenic systems. In the present work we use one of the representations of the Coulomb Green's function to calculate the generalized dynamic polarizability. The use of Coulomb units enables us to easily extend the results obtained to the case of hydrogenic ions. Other methods, such as the Sternheimer method,[11] can be used to obtain the generalized polarizability. We used the Coulomb Green's function in the coordinate representation, but similar calculations can be performed using other representations of the Green's function, for example, the momentum representation.[12]

The proposed method for calculating the generalized dynamic polarizability of hydrogen can, in principle, yield closed analytic expressions for the generalized polarizability of a state with any principal quantum number. The idea of the method is to represent the radial wave functions of the hydrogen atom R_{nl} by differentiating a generating function of Laguerre polynomials:[9]

$$R_{nl}(r)=\frac{2^{l+1}}{n^{l+2}\sqrt{\Gamma(n-l)\Gamma(n+l+1)}}\times\frac{d^{n-l-1}}{dt^{n-l-1}}\left[\frac{r^le^{-\lambda r}}{(1-t)^{2l+2}}\right]\Bigg|_{t=0}, \quad (5)$$

where $\lambda=1/n\times(1+t)/(1-t)$.

Such a representation of wave functions is convenient for calculating matrix elements and, thus, the polarizability (2), since it permits performing the differentiation operation after calculating the radial integral in the matrix elements and thereby significantly simplifies the entire calculation. Actually, matrix elements which contain $r^le^{-\lambda r}$ instead of the radial wave functions of the hydrogen atom must be calculated. The final result for various states is then obtained by differentiating the expression obtained different numbers of times.

This method is especially convenient if the differentiation with respect to a parameter is carried out by a programmed procedure (we used the MATHEMATICA software from Wolfram Research).

We now show that the sum of the matrix elements shaping the polarizability can be expressed in terms of a set of hypergeometric functions. Using the dispersion representation of a Green's function, we write the polarizability in the form of a sum of two terms

$$\alpha_j(\omega,q)=X(E_j+\omega,q)+X(E_j-\omega,q), \quad (6)$$

where

$$X(E,q)=-\frac{i}{q}\langle j|e^{-i\mathbf{q}\cdot\mathbf{r}_1}G(\mathbf{r}_1,\mathbf{r}_2;E)\mathbf{r}_2\cdot\mathbf{n_q}|j\rangle.$$

We next utilize the expansion of the Green's function in spherical waves:

$$G(\mathbf{r}_1,\mathbf{r}_2;E)=\sum_{lm}G_l(r_1,r_2;E)Y_{lm}(\mathbf{n}_1)Y^*_{lm}(\mathbf{n}_2), \quad (7)$$

where for the Green's function of the radial Sternheimer equation we use the representation[13]

$$G_l(r_1,r_2;E)=\frac{2i}{\sqrt{r_1r_2}}(-1)^{l+1}\int_1^\infty\frac{d\xi}{\sqrt{\xi^2-1}}\left(\frac{\xi+1}{\xi-1}\right)^{i\nu}\times J_{2l+1}(2k\sqrt{r_1r_2}\sqrt{\xi^2-1})e^{ik\xi(r_1+r_2)},$$

$$k=\sqrt{2E}, \quad \nu=\frac{1}{k}. \quad (8)$$

This representation is formally valid for $E>0$. However, it can be continued analytically into the region $E<0$. We perform analytical continuation in the final expressions after expressing $X(E,q)$ in terms of hypergeometric functions.

Let us consider the case of $l=0$, which corresponds to the selection of s states of the hydrogen atom in the initial

and final states of the process, in greater detail. In this case, after integration over the angles, the matrix element $X(E,q)$ can be written in the following form:

$$X(E,q)=\frac{4}{n^5\Gamma^2(n)}\hat{D}_1\hat{D}_2M. \tag{9}$$

Here the differentiation operators are defined by the expression

$$\hat{D}_m f(t_m)=\frac{d^{n-1}}{dt_m^{n-1}}\left[\frac{f(t_m)}{(1-t_m)^2}\right]\Bigg|_{t_m=0}, \quad m=1,2, \tag{10}$$

and the radial integral can be represented in the following form after integration over r_1 and r_2 and replacement of the variable:

$$M=2^9ik^3\int_0^1 d\tau(1-\tau)\tau^{1-i\nu}\frac{by-ax\tau}{[(by-ax\tau)^2+q^2(y-x\tau)^2]^3}$$
$$-2^9ik^5\int_0^1 d\tau\tau^{2-i\nu}\frac{5(by-ax\tau)^2-q^2(y-x\tau)^2}{[(by-ax\tau)^2+q^2(y-x\tau)^2]^4}, \tag{11}$$

where $a=\lambda_1+ik$, $b=\lambda_1-ik$, $x=\lambda_2+ik$, and $y=\lambda_2-ik$.

If $q=0$, the integral (11) reduces to the definition of a hypergeometric function. This case corresponds to the limiting transition of the generalized polarizability to the dynamic analog. As a result, we obtain a fairly simple expression for $X(E,0)$ in terms of two hypergeometric functions

$$X(E,0)=\frac{2^{11}}{n^5\Gamma^2(n)}\hat{D}_1\hat{D}_2\left[\frac{ik^3}{b^5y^5(2-i\nu)(3-i\nu)^2}\right.$$
$$\times F_1\left(5,2-i\nu;4-i\nu;\frac{ax}{by}\right)-\frac{5ik^5}{b^6y^6(3-i\nu)^2}$$
$$\left.\times F_1\left(6,2-i\nu;4-i\nu;\frac{ax}{by}\right)\right]. \tag{12}$$

The dynamic dipole polarizability is obtained by substituting (12) into (6).

If $q\neq 0$, the integral (11) can also be represented in terms of hypergeometric functions by expanding the integrands in partial fractions. However, the expression obtained is fairly cumbersome; therefore, its explicit form is presented in the Appendix.

FEATURES OF THE GENERALIZED DYNAMIC DIPOLE POLARIZABILITY OF EXCITED STATES OF HYDROGEN ATOMS

Let us analyze formula (2), calling the first term in the square brackets term A and the second term B. We first consider the frequency dependence $\alpha_j(\omega,q)$. For the intermediate states $|f\rangle$ in the discrete spectrum we have $E_f<0$. If E_f is greater than the energy of the initial and final states E_j, i.e., if $\omega_{fj}>0$, term B behaves monotonically as the frequency ω increases, while term A exhibits a resonance increase near the pole $\omega\to E_f-E_j$. Polarizability poles correspond to a real process, in which, as the hydrogen atom interacts with the projectile, it is excited to the intermediate state $|f\rangle$ and then returns to the initial state $|j\rangle$ with the emission of a photon. The divergence of the polarizability is eliminated by introducing the imaginary addition $i\Gamma_i$, where Γ_i is the width of the level, into the energy denominator. In this paper we shall consider only values of ω for which the condition $\omega-(E_f-E_j)\gg\Gamma_i$ holds. Because the radiation width Γ_i is small, this condition holds up to frequencies very close to the polarizability poles.

If the initial state $|j\rangle$ of the hydrogen atom is the ground state, some intermediate states $|f\rangle$ satisfy the inequality $0>E_j>E_f$, i.e., $\omega_{fj}<0$. When this condition holds, term A behaves monotonically as ω increases. Conversely, term B undergoes a resonance increase at $\omega\to E_j-E_f$. This divergence also corresponds to a real process, in which a hydrogen atom initially in an excited state $|j\rangle$ is de-excited with the emission of a photon and then returns to the original excited state when it interacts with a projectile. In this case the excited state of the hydrogen atom satisfies the condition $I\equiv -E_j<E_j-E_f$, i.e., the polarizability pole $\omega=E_j-E_f$ is above the ionization threshold I of the respective excited state $|j\rangle$.

Thus, the frequency dependence of the polarizability has poles, which correspond to real excitation and de-excitation processes of an atom in an intermediate state. Also, the poles corresponding to de-excitation processes lie above the ionization potentials of the corresponding excited states. Therefore, the polarization bremsstrahlung spectrum generated during a collision with an excited hydrogen atom should display narrow lines not only below, but also above the ionization potential of the excited state under consideration. The only exception is the $2s$ state, since dipole transitions with de-excitation are forbidden for this state by the selection rules.

Figure 1(a) shows the polarizability of the $3s$ state of the hydrogen atom near the frequency corresponding to the $3s\to 2p$ transition. It is seen that the real part of the polarizability has a clearly expressed pole character. The imaginary part is described only by term A and decreases monotonically with increasing ω.

Let us move on to an analysis of the contribution of intermediate states with an energy $E_f>0$ to $\alpha(\omega,q)$. According to the asymptotic form of the wave functions of the intermediate state, the product of two rapidly oscillating functions of the types e^{-iqr} and e^{ikr}, where k is the wave vector of the electron in the intermediate state, appears in the matrix element $\langle j|e^{-i\mathbf{q}\cdot\mathbf{r}}|f\rangle$ at fairly large values of E_f and q. When $q=k$, i.e., when $E_f=q^2/2$, the oscillations of the functions in the integrand are mutually compensated, and, as a result, the matrix element has a maximum as a function of q in the vicinity of the point $q=k$. When $\alpha(\omega,q)$ is calculated and the integration over the energy of the intermediate state E_f is carried out, the states with $E_f\sim q^2/2$ make the main contribution. Due to the small value of the energy denominator, this leads to the appearance of a feature in the generalized dynamic polarizability at

$$\omega\approx\frac{q^2}{2}+I. \tag{13}$$

The region of mutual compensation of the exponential functions e^{-iqr} and e^{ikr} is determined by the radius of the initial state. As the radius of the state increases, the feature

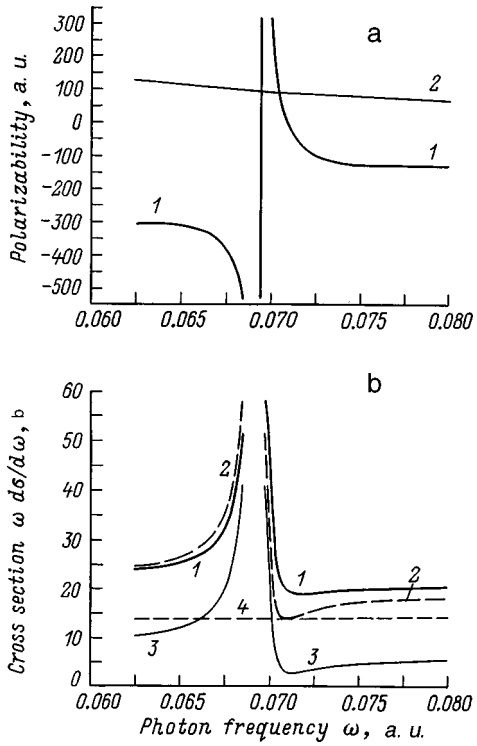

FIG. 1. a—Dependence of the generalized dynamic polarizability of the $3s$ state of the hydrogen atom on the photon frequency ω near the polarizability pole at $\omega = E_{3s} - E_{2p} = 5/72$ for a fixed momentum transfer $q = 0.001$: *1*—real part of the polarizability, *2*— imaginary part of the polarizability, vertical line—position of the pole of the real part of the polarizability. b—Spectral dependence of the bremsstrahlung cross section for electrons and positrons impinging on hydrogen atoms in the $3s$ state and a projectile velocity $v = 4$: *1* (heavy line)—total bremsstrahlung cross section for electrons, *2*—total bremsstrahlung cross section for positrons, *3*—polarization bremsstrahlung cross section, *4*—ordinary bremsstrahlung cross section.

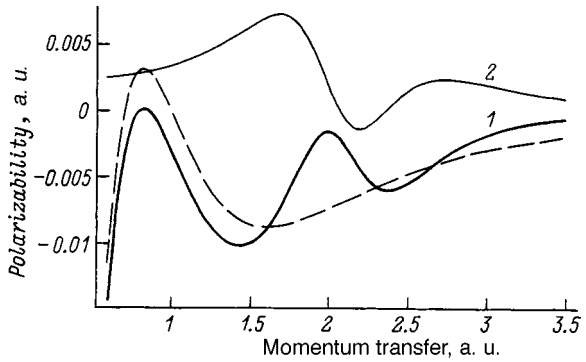

FIG. 2. Dependence of the generalized dynamic polarizability of the $2s$ state of the hydrogen atom on the momentum transfer q at a fixed photon frequency $\omega = 2.14$. *1*—real part of the polarizability, *2*—imaginary part of the polarizability, dashed curve—asymptotic form of the real part of the polarizability $\alpha(\omega,q) \propto -W(q)/\omega^2$.

(13) becomes more pronounced. This conclusion is confirmed by our calculations. The feature (13) is scarcely noticeable in the case of the $1s$ state and intensifies upon transition to the $2s$ and $3s$ states.

Let us now consider the dependence of the generalized polarizability on the momentum transfer q. An important feature of the excited states of the hydrogen atom is the considerable inhomogeneity of the spatial distribution of the electron charge. This leads to the appearance of zeros in the Fourier transform of the charge density, i.e., in the form factor. It can be shown that the form factor of the s states of the hydrogen atom is described by a rational function of q^2 and that this function does not have zeros only in the case of the $1s$ state. In the limit of large photon frequencies, $\omega \gg I$, the polarizability of an atom is proportional to its form factor: $\alpha_j(\omega,q) \propto -W_j(q)/\omega^2$. Therefore, the nonmonotonic behavior of the form factor as a function of q also leads to nonmonotonic behavior of the generalized polarizability as a function of q.

Figure 2 illustrates the dependence of the polarizability of the $2s$ state on the momentum transfer q. The solid heavy line shows the dependence of the real part of the generalized dynamic polarizability on the momentum transfer q at a fixed frequency $\omega = 2.14$. The real part of the polarizability has two maxima in the interval $q \in [0.5,3.5]$. The first appears because of the nonmonotonic behavior of the form factor. The form factor of the $2s$ state divided by minus the square of the frequency is represented by the dashed curve. The second maximum is a manifestation of the feature described above, which appears at $\omega \approx q^2/2 + I$. Curve *2* shows the behavior of the imaginary part of the dynamic polarizability. The feature at the point (13) is also manifested in the irregular behavior of the imaginary part of the polarizability as a function of q.

RESULTS OF A NUMERICAL CALCULATION OF BREMSSTRAHLUNG CROSS SECTIONS

The results of a numerical calculation of the total cross sections of the bremsstrahlung appearing when light charged particles (electrons and positrons) are scattered on an excited hydrogen atom are shown in Figs. 1, 3, and 4.

Figure 1(b) presents the spectral dependence of the cross section of the bremsstrahlung formed during collisions of electrons and positrons with a hydrogen atom in the $3s$ state. The range of photon frequencies contains the polarizability pole corresponding to the $3s \rightarrow 2p$ transition. The energy of this transition is greater than the ionization potential of the $3s$ subshell. Thus, the fact that the generalized polarizability has a pole above the ionization potential leads to the appearance of a narrow line in the bremsstrahlung spectrum. We note that the interference term in the total cross section of the process is proportional to the real part of the polarizability, which changes sign as it passes the pole. The change in the sign of the interference between the ordinary and polarization bremsstrahlung mechanisms is manifested by the fact that below the polarizability pole the bremsstrahlung cross section for positrons is greater than the cross section for electrons, and, conversely, above the pole the bremsstrahlung cross section for electrons is greater than the cross section for positrons.

Figure 3 presents the spectral dependences of the bremsstrahlung cross section for the $1s$, $2s$, and $3s$ states of the hydrogen atom when the projectile velocity $v = 4$ a.u. A comparison of these plots reveals that at a fixed collision velocity the contribution of the polarization bremsstrahlung mechanism to the total cross section of the process falls off

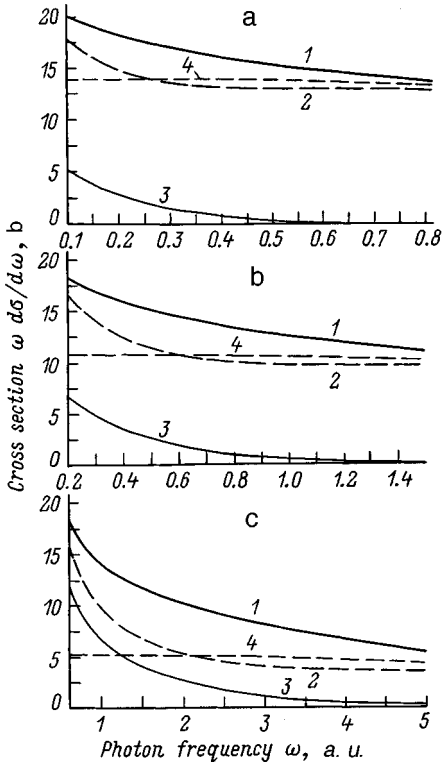

FIG. 3. Spectral dependence of the bremsstrahlung cross sections for electrons and positrons colliding with hydrogen atoms in the $1s$ (c), $2s$ (b), and $3s$ (a) states and a projectile velocity $v=4$: *1*—total bremsstrahlung cross section for electrons, *2*—total bremsstrahlung cross section for positrons, *3*—polarization bremsstrahlung cross section, *4*—ordinary bremsstrahlung cross section.

as the principal quantum number rises. While the polarization bremsstrahlung mechanism is more efficient near the threshold in the $1s$ state, the ordinary bremsstrahlung cross section exceeds the polarization bremsstrahlung cross section in the $2s$ and $3s$ states. This is associated with the increase in the orbital radius as the principal quantum number increases. In fact, small values of the impact parameter, at which the projectile travels near the nucleus of the target atom, are important for the formation of bremsstrahlung by the ordinary mechanism. Therefore, an increase in the radius of the target atom leads to an increase in the ordinary bremsstrahlung cross section. Conversely, bremsstrahlung is formed most efficiently according to the polarization mechanism at large values of the impact parameter, where the projectile strongly polarizes the target atom.

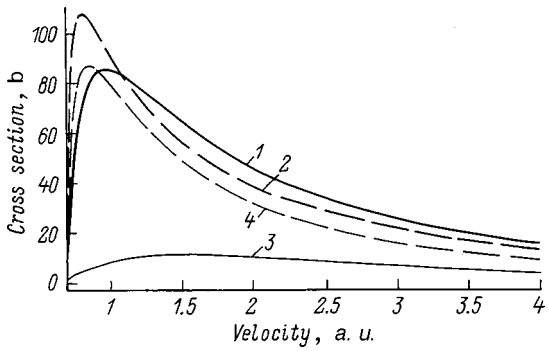

FIG. 4. Dependence of the bremsstrahlung cross sections for electrons and positrons on projectile velocity for hydrogen atoms in the $2s$ state and a photon frequency $\omega=0.25$: *1*—total bremsstrahlung cross section for electrons, *2*—total bremsstrahlung cross section for positrons, *3*—polarization bremsstrahlung cross section, *4*—ordinary bremsstrahlung cross section.

Figure 4 shows the dependence of the cross section of the process on the projectile velocity at $\omega=0.25$ a.u., which amounts to twice the ionization potential of the $2s$ subshell. This dependence has a characteristic maximum, and consideration of the interference between the ordinary and polarization bremsstrahlung mechanisms leads to different positions for the maxima of the cross section of the process for electrons and positrons. It is also noteworthy that the interference between the two bremsstrahlung mechanisms has a constructive character for electrons and a destructive character for positrons at $v>1.1$, while electrons and positrons change roles at $v<1.1$. Such behavior of the cross section is a result of the sign-alternating character of $\alpha(\omega,q)$ as a function of q [see Eq. (3)], which, in turn, is a result of the diffraction of electrons on the inhomogeneous charge distribution in the $2s$ state of the hydrogen atom.

CONCLUSION

Thus, a method for obtaining closed analytic expressions for the generalized dynamic dipole polarizability of the hydrogen atom and hydrogenic ions in both the ground and excited states has been proposed in this paper. The method is based on the use of the Coulomb Green's function and the simultaneous representation of the wave functions of the initial and final states of an electron in terms of a generating function of Laguerre polynomials.

The proposed method has been used to calculate the generalized polarizability of the $1s$, $2s$, and $3s$ states. An analysis of the expressions obtained showed that the polarizability of the excited state has features not observed for the polarizability of the ground state.

Virtual transitions with de-excitation in an intermediate state are possible for excited states. The energies of such transitions are greater than the ionization potential of the initial excited state. This leads to the appearance of narrow lines in the continuous emission spectrum with energies above the ionization potential of the atom.

Another important feature of the excited states is the nonmonotonic dependence of the differential bremsstrahlung cross sections on the momentum transfer q. This feature has been associated with the inhomogeneity of the electron charge distribution in the excited states.

Our calculation shows that the local maximum in the differential bremsstrahlung cross section at $\omega\sim q^2/2$ is more pronounced for the excited $2s$ and $3s$ states than for the $1s$ state. This maximum is a manifestation of the familiar Bethe feature (the Bethe ridge) in the differential inelastic scattering of electrons on atoms.

This work was made possible by the financial support of the Russian Fund for Fundamental Research (Project No. 96-02-17922-a) and the International Scientific-Technical Center (Project No. 076-95).

APPENDIX

The explicit form of the expression for calculating the generalized dynamic polarizability of a hydrogen atom in a state with a principal quantum number equal to n and an angular momentum quantum number $l=0$ is presented below.

The polarizability is expressed in terms of the sum of two terms containing differentiation of the hypergeometric functions ${}_2F_1$. We note that the resultant expression can be transformed into a more compact form by reducing the number of hypergeometric functions. However, in our opinion, the representation given is more convenient from the standpoint of performing numerical calculations

$$\alpha_{ns}(\omega,q)=X(E_n+\omega,q)+X(E_n-\omega,q), \tag{A1}$$

where

$$X(E,q)=\frac{2^6}{n^5\Gamma^2(n)}\hat{D}_1\hat{D}_2M. \tag{A2}$$

The differential operators $\hat{D}$ are defined by (10), and M has the following form:

$$\begin{aligned}
M=&-3\frac{\chi^2\alpha}{y^5\gamma q^3k(2-i\nu)(3-i\nu)}{}_2F_1(1,2-i\nu,4-i\nu,z_1)\\
&+3\frac{\chi\alpha^2}{y^5\beta q^3k(2-i\nu)(3-i\nu)}{}_2F_1(1,2-i\nu,4-i\nu,z_2)\\
&+6i\frac{\chi^2\alpha}{y^6\gamma q^3(3-i\nu)}{}_2F_1(1,3-i\nu,4-i\nu,z_1)\\
&-6i\frac{\chi\alpha^2}{y^6\beta q^3(3-i\nu)}{}_2F_1(1,3-i\nu,4-i\nu,z_2)\\
&+2i\frac{\chi^2(\alpha+2iq)}{y^5\gamma^2q^3k(2-i\nu)(3-i\nu)}{}_2F_1(2,2-i\nu,4-i\nu,z_1)\\
&-2i\frac{\alpha^2(\chi-2iq)}{y^5\beta^2q^3(2-i\nu)(3-i\nu)}{}_2F_1(2,2-i\nu,4-i\nu,z_2)\\
&+6\frac{\chi^2k(\alpha+iq)}{y^6\gamma^2q^3(3-i\nu)}{}_2F_1(2,3-i\nu,4-i\nu,z_1)\\
&-6\frac{\alpha^2(\chi-iq)}{y^6\beta^2q^3(3-i\nu)}{}_2F_1(2,3-i\nu,4-i\nu,z_2)\\
&+8i\frac{\chi^2k}{y^5\gamma^3q^2(2-i\nu)(3-i\nu)}{}_2F_1(3,2-i\nu,4-i\nu,z_1)\\
&+8i\frac{\alpha^2k}{y^5\beta^3q^2(2-i\nu)(3-i\nu)}{}_2F_1(3,2-i\nu,4-i\nu,z_2)\\
&-4i\frac{\chi^2k^2(\alpha+4iq)}{y^6\gamma^3q^3(3-i\nu)}{}_2F_1(3,3-i\nu,4-i\nu,z_1)\\
&+4i\frac{\alpha^2k^2(\chi-4iq)}{y^6\beta^3q^3(3-i\nu)}{}_2F_1(3,3-i\nu,4-i\nu,z_2)\\
&-24i\frac{\chi^2k^3}{y^6\gamma^4q^2(3-i\nu)}{}_2F_1(4,3-i\nu,4-i\nu,z_1)\\
&-24i\frac{\alpha^2k^3}{y^6\beta^4q^2(3-i\nu)}{}_2F_1(4,3-i\nu,4-i\nu,z_2).
\end{aligned} \tag{A3}$$

The following notation was used in Eq. (A3):

$$\begin{aligned}
&k=\sqrt{2E},\quad \nu=\frac{1}{k},\\
&z_1=\frac{\chi x}{\gamma y},\quad z_2=\frac{\alpha x}{\beta y},\\
&\alpha=\lambda_1+ik+iq,\quad \beta=\lambda_1-ik+iq,\\
&\chi=\lambda_2+ik-iq,\quad \gamma=\lambda_2-ik-iq,\\
&x=\lambda_2+ik,\quad y=\lambda_2-ik,\\
&\lambda_1=\frac{1}{n}\frac{1+t_1}{1-t_1},\quad \lambda_2=\frac{1}{n}\frac{1+t_2}{1-t_2}.
\end{aligned} \tag{A4}$$

[1] A. I. Akhiezer and V. B. Berestetskii, *Quantum Electrodynamics* [Wiley, New York (1965); 4th ed., Nauka, Moscow (1981), 432 pp.].
[2] R. H. Pratt, in *Fundamental Processes in Energetic Atomic Collisions*, edited by H. O. Lutz, J. S. Briggs, and H. Kleinpoppen [Plenum Press, New York (1983)], pp. 145–182.
[3] V. N. Tsytovich and I. M. Oiringel (eds.), *Polarization Bremsstrahlung* [Plenum Press, New York (1992); Nauka, Moscow (1987), 335 pp.].
[4] M. Ya. Amusia, Phys. Rep. **162**, 145 (1988).
[5] A. V. Korol and A. V. Solov'yov, J. Phys. B **30**, 1105 (1997).
[6] V. M. Buĭmistrov and L. I. Trakhtenberg, Zh. Eksp. Teor. Fiz. **69**, 108 (1975) [Sov. Phys. JETP **42**, 54 (1976)].
[7] V. M. Buĭmistrov, Yu. A. Krotov, and L. I. Trakhtenberg, Zh. Éksp. Teor. Fiz. **79**, 808 (1980) [Sov. Phys. JETP **52**, 411 (1980)].
[8] A. Duboius and A. Maquet, Phys. Rev. A **40**, 4288 (1989).
[9] V. A. Fock, *The Principles of Quantum Mechanics* [Mir, Moscow (1978); 2nd ed., Nauka, Moscow (1976), 376 pp.].
[10] A. V. Korol, A. G. Lyalin, and A. V. Solov'yov, J. Phys. B **30**, L115 (1997).
[11] V. Yakhontov and K. Jungmann, Z. Phys. D At. Mol. Clusters **38**(2), 141 (1996).
[12] V. G. Gorshkov, Zh. Éksp. Teor. Fiz. **47**, 352 (1964) [Sov. Phys. JETP **20**, 234 (1965)].
[13] B. A. Zon, N. L. Manakov, and L. P. Rapoport, *Theory of Multiphoton Processes in Atoms* [in Russian], Atomizdat, Moscow (1978), 184 pp.

Translated by P. Shelnitz

Instability and breakup of a thick layer of a viscous magnetic fluid in a tilted magnetic field

V. M. Korovin

Scientific-Research Institute of Mechanics, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted May 29, 1998)
Zh. Tekh. Fiz. **69**, 14–22 (October 1999)

A linear partial differential equation describing the evolution of an initial disturbance of a flat free surface of a thin layer of a viscous magnetic fluid covering a horizontal plate in the presence of a uniform magnetic field, is derived within a system of ferrohydrodynamic and magnetostatic equations. The effect of magnetizing the plate on the stability of the flat free surface is investigated. An estimate is obtained for the minimum value of the tangential component of the magnetization vector of the fluid sufficient to radically alter the pattern of the final breakup of the continuous layer. © *1999 American Institute of Physics.* [S1063-7842(99)00310-4]

INTRODUCTION

It is known from theoretical studies and experiments[1–4] that a fairly strong uniform magnetic field perpendicular to a flat free surface of a motionless magnetic fluid (or a flat interface between immiscible fluids with different magnetic permeabilities) causes instability. As a result, the fluid passes from a hydrostatic state with a flat free surface to another hydrostatic state. In the case of thick layers, a periodic surface structure consisting of hexagonal cells forms, whereas a thin layer on a solid substrate breaks up into separate drops.[3,4]

The theoretical analysis in Ref. 4 of the behavior of a stable stratified (in the absence of a field) configuration consisting of two layers of magnetic fluids lying one on top of the other and immersed in a tilted magnetic field **H**, whose normal component exceeds the critical value, showed that the tangential component $\mathbf{H}_\tau$ stabilizes a certain range of harmonic disturbances propagating along $\mathbf{H}_\tau$. On the other hand, the tangential component does not have any influence on disturbances propagating transversely to $\mathbf{H}_\tau$. It was concluded on this basis in Ref. 4 that a corrugated interface with corrugation axes parallel to $\mathbf{H}_\tau$ forms, but an estimate of the magnitude of $\mathbf{H}_\tau$ needed for the actual realization of such a scenario for the development of instability of a flat interface was not obtained in Ref. 4.

In the previous studies the instability of a flat free surface of a magnetic fluid caused by an orthogonal magnetic field was investigated within the theory of plane-parallel potential motions of an inviscid fluid. It should be noted in this context that in the closely related hydrodynamic problem of the damping of small-amplitude gravity waves[5] (as well as capillary gravity waves[6]) on a free surface of a nonmagnetic Newtonian fluid of infinite depth (for $kd\to\infty$, where k is the wave number and d is the depth) the model of an ideal fluid is applicable only in the case of $\lambda^2|\omega|\gg\nu$, where $\lambda=2\pi/k$ is the wave number, ω is the frequency, and ν is the kinematic viscosity of the fluid under consideration. If $d/\lambda\sim 1$ and $|\omega|\tau_d\sim 1$, where $\tau_d=d^2/\nu$ is the characteristic time for the diffusion of vorticity across the fluid layer, the damping of the waves due to the viscosity of the fluid is significant. Of course, the motion is turbulent.

In the general case the frequency $\omega(k)$ is a complex-valued function of the real wave number; however, for any $\nu>0$ there is a limiting layer thickness d_m, which is such that if $d\leqslant d_m$ the frequency becomes purely imaginary, so that if $d\leqslant d_m$ no sinusoidal progressive waves exist.[6] In addition, the value of d_m increases with increasing ν.

Unlike the previous studies, the present work explores the instability and breakup of a thin $(d/\lambda\ll 1)$ layer of a Newtonian magnetic fluid caused by a tilted magnetic field in the case of $\tau_d/\tau_i\ll 1$, where τ_i is the characteristic time for the development of instability.

INITIAL EQUATIONS AND FORMULATION OF THE PROBLEM

A layer of a motionless magnetic fluid with a flat free surface on a horizontal nonmagnetic plate immersed in a uniform tilted magnetic field is considered (Fig. 1). We assume that the fluid is homogeneous in composition and that its permeability μ depends only on the amplitude of the field.

We introduce the Cartesian coordinate system x_1, x_2, x_3, whose x_1 axis is directed along the horizontal component of the field, while the x_3 axis points vertically upward and the $x_3=0$ plane coincides with the free surface of the fluid. Let f_0 be the potential of the field in the region with flat boundaries occupied by the magnetic fluid, and let f_{01} and f_{02} be the potentials of the fields in the regions lying above and below that layer, respectively. We use d to denote the thickness of the layer. In the coordinate system under consideration we have

$$f_0=H_0(x_1\sin\vartheta+x_3\cos\vartheta),$$
$$f_{01}=H_{a0}(x_1\sin\vartheta_a+x_3\cos\vartheta_a),$$
$$f_{02}=f_{01}-d(H_0\cos\vartheta-H_{a0}\cos\vartheta_a), \qquad (1)$$

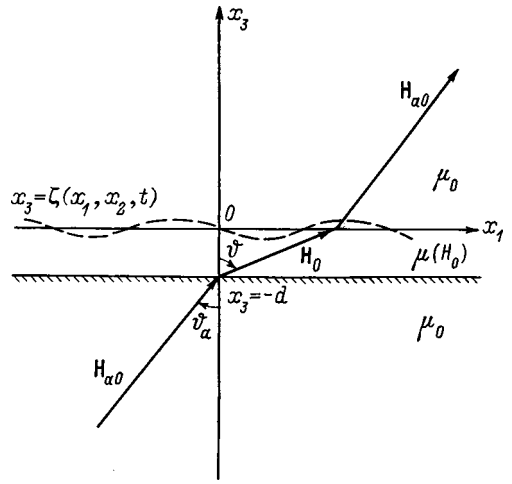

FIG. 1. Geometry of the problem.

where the field amplitudes H_0 and H_{a0} within the layer of magnetic fluid and outside it, respectively, are shown in Fig. 1, the tilt angles ϑ and ϑ_a of $\mathbf{H}_0$ and $\mathbf{H}_{a0}$ relative to the x_3 axis are related by the conditions known from magnetostatics on an interface between media with different magnetic properties

$$H_{a0}\sin\vartheta_a = H_0\sin\vartheta,$$

$$\mu_0 H_{a0}\cos\vartheta_a = \mu(H_0)H_0\cos\vartheta, \tag{2}$$

and $\mu_0 = 4\pi\times 10^{-7}$ H/m is the permeability of free space.

In addition, H_{a0} and ϑ_a in (1) are assumed to be assigned, and the equalities (2) are used to calculate H_0 and ϑ.

The induction within the flat layer $\mathbf{B}_0=\mu(H_0)\mathbf{H}_0$ and the magnetization of the fluid $\mathbf{M}_0=\mathbf{B}_0/\mu_0-\mathbf{H}_0$ are homogeneous, and the pressure increases linearly with increasing distance from the free surface according to the law $P_0=p_a-\rho g x_3-\mu_0 M_{03}^2/2$, where p_a is the air pressure near the free surface, ρ is the density of the fluid, and g is the acceleration of gravity. The last term in the expression for P_0 represents the magnetic pressure jump.[4]

Let us formulate the linear problem of the stability of the hydrostatic state P_0, $\mathbf{M}_0$ of a fluid layer with a flat free surface. Let $x_3=\zeta(x_1,x_2,t)$ represent the equation of the free surface in the presence of small-amplitude waves, where t denotes the time. Introducing the potentials of small disturbances of the magnetic fields within $[f(x_1,x_2,x_3,t)]$ and outside $[f_1(x_1,x_2,x_3,t)]$ the fluid and $j=1,2$, we write the magnetic fields, as well as the induction and magnetization of the fluid, disturbed by the wave-induced deformation of the free surface in the form

$$\mathbf{H}=\nabla(f_0+f), \quad \mathbf{H}_j=\nabla(f_{0j}+f_j), \quad j=1,2,$$

$$\mathbf{B}=\mathbf{B}_0+\mathbf{b}, \quad \mathbf{M}=\mathbf{M}_0+\mathbf{m}.$$

We can similarly represent the distribution of the pressure P in the fluid in the presence of waves: $P=P_0+p(x_1,x_2,x_3,t)$. To within small first-order quantities we have

$$H-H_0=\frac{\partial f}{\partial x_1}\sin\vartheta+\frac{\partial f}{\partial x_3}\cos\vartheta,$$

$$M-M_0=m_1\sin\vartheta+m_3\cos\vartheta,$$

$$\mathbf{m}=\mathbf{a}_1\left[\frac{\partial f}{\partial x_1}(\chi_0\cos^2\vartheta+\chi_t^0\sin^2\vartheta)+\frac{1}{2}\frac{\partial f}{\partial x_3}\times(\chi_t^0-\chi_0)\sin 2\vartheta\right]+\mathbf{a}_2\chi_0\frac{\partial f}{\partial x_2}+\mathbf{a}_3\left[\frac{1}{2}\frac{\partial f}{\partial x_1}\times(\chi_t^0-\chi_0)\sin 2\vartheta+\frac{\partial f}{\partial x_3}(\chi_0\sin^2\vartheta+\chi_t^0\cos^2\vartheta)\right],$$

$$\mathbf{b}=\mu_0(\mathbf{m}+\nabla f), \quad \chi_0=\frac{M_0}{H_0}, \quad \chi_t^0=\left.\frac{dM}{dH}\right|_{H=H_0},$$

$$\mu_r^0=1+\chi_0, \quad \mu_{rt}^0=1+\chi_t^0, \tag{3}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are unit vectors along the corresponding coordinate axes.

We confine ourselves to an investigation of the instability of a thin layer, where $\varepsilon=d/\lambda\ll 1$. In this case, neglecting the terms in orders of ε (which appear because of the nonlinearity of the magnetization law) and higher terms in the equation $\operatorname{div}\mathbf{b}=0$ written relative to f, we have

$$\frac{\partial^2 f}{\partial x_3^2}=0, \quad \Delta f_j=0, \quad j=1,2,$$

$$\Delta=\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}+\frac{\partial^2}{\partial x_3^2}. \tag{4}$$

With consideration of the first expression in (3), the linearized equation of motion of a Newtonian magnetic fluid takes the form

$$\rho\frac{\partial\mathbf{v}}{\partial t}=-\nabla\rho+\eta\Delta\mathbf{v}+\mu_0 M_0\nabla\left(\frac{\partial f}{\partial x_1}\sin\vartheta+\frac{\partial f}{\partial x_3}\cos\vartheta\right), \tag{5}$$

where $\mathbf{v}=(v_1,v_2,v_3)$ is the velocity and η is the dynamic viscosity.

We note that if the fluid forming the thin layer undergoes wave motion, the condition for linearization of the equation of motion is $\zeta_*/d\ll 1$, where ζ_* is the characteristic amplitude of the waves. This condition is assumed to be satisfied below.

Equation (5) can be simplified in the case under consideration: to the accuracy adopted in writing the first equality in (4), the replacement of $\Delta\mathbf{v}$ by $\partial^2\mathbf{v}/\partial x_3^2$ is permissible in the equation of motion (5). Then, in estimating the ratio of the inertial term in Eq. (5) to the viscous term to an order of magnitude, we have

$$\rho\frac{\partial v_j}{\partial t}\left(\eta\frac{\partial^2 v_j}{\partial x_3^2}\right)^{-1}\sim\frac{\tau_d}{\tau_i}, \quad j=1,2,3;$$

$$\tau_d=\frac{d^2}{\nu}; \quad \nu=\frac{\eta}{\rho}. \tag{6}$$

In the case of the appearance of instability, τ_i is the characteristic time for the development of the most rapidly growing harmonic [see the corresponding formula (21)]. If

the flat free surface of the magnetic fluid is stable for an assigned $\mathbf{H}_{a0}$, τ_i is the characteristic damping time of the disturbances.

We shall henceforth consider the case of $\tau_d/\tau_i \ll 1$. Since the characteristic time for the diffusion of vorticity across the layer τ_d (6) falls off with increasing ν, it is natural to expect that the condition $\tau_d/\tau_i \ll 1$ holds for not excessively thick layers of fairly viscous fluids. The motion of the magnetic fluid can then be described by a system consisting of the equations

$$\operatorname{div}\mathbf{v}=0,$$

$$\eta\frac{\partial^2\mathbf{v}}{\partial x_3^2}=\nabla\left[p-\mu_0 M_0\left(\frac{\partial f}{\partial x_1}\sin\vartheta+\frac{\partial f}{\partial x_3}\cos\vartheta\right)\right] \tag{7}$$

and Eq. (4). The condition for applicability of the equation of motion in the form (7) is written out below in the form of Eq. (22).

The kinetic and dynamic conditions, as well as the condition of continuity of the tangential component of the magnetic field and the normal component of the induction on the free surface, are written in the approximation under consideration in the following manner:

$$x_3=0:\ \frac{\partial\zeta}{\partial t}=v_3,\quad \frac{\partial v_1}{\partial x_3}=0,\quad \frac{\partial v_2}{\partial x_3}=0,$$

$$p=\rho g\zeta+2\eta\frac{\partial v_3}{\partial x_3}-\alpha\left(\frac{\partial^2\zeta}{\partial x_1^2}+\frac{\partial^2\zeta}{\partial x_2^2}\right)$$

$$+\mu_0 M_0\cos\vartheta\left[M_0\sin\vartheta\frac{\partial\zeta}{\partial x_1}-\gamma\frac{\partial f}{\partial x_3}\right],$$

$$f-f_1=M_0\zeta\cos\vartheta,$$

$$(1+\gamma)\frac{\partial f}{\partial x_3}-\frac{\partial f_1}{\partial x_3}=M_0\sin\vartheta\frac{\partial\zeta}{\partial x_1}, \tag{8}$$

where $\gamma=\chi_0\sin^2\vartheta+\chi_t^0\cos^2\vartheta$.

The boundary conditions on the surface of the nonmagnetic plate have the form

$$x_3=-d:\ v_1=0,\quad v_2=0,\quad v_3=0,$$

$$f=f_2,\quad (1+\gamma)\frac{\partial f}{\partial x_3}=\frac{\partial f_2}{\partial x_3}. \tag{9}$$

The disturbances of the magnetic field, of course, vanish as the distance between the point of observation and the free surface of the fluid increases:

$$\nabla f_1\to 0\ \text{ as }\ x_3\to\infty,\quad \nabla f_2\to 0\ \text{ as }\ x_3\to-\infty. \tag{10}$$

The formulation of the problem of the instability of the hydrostatic state of a fluid layer with a flat free surface [Eqs. (4) and (7)–(10)] also includes assignment of an initial condition for the function sought $\zeta(x_1,x_2,t)$.

EVOLUTION OF AN INITIAL DISTURBANCE OF THE FREE SURFACE SHAPE AND CALCULATION OF THE VELOCITY FIELD AND THE PRESSURE DISTURBANCE

To simplify the ensuing mathematical manipulations, we reduce the system of hydrodynamic equations (7) with the corresponding boundary conditions (8) and (9) to a single equation in partial derivatives, which relates $\zeta(x_1,x_2,t)$ to $f(x_1,x_2,x_3,t)$. For this purpose, as a first step we integrate the projection of the simplified pressure equation (7) onto the x_3 axis. With consideration of the first equality in (4), from which it follows that $\partial f/\partial x_3$ does not depend on x_3, we obtain

$$p=\eta\frac{\partial v_3}{\partial x_3}+\mu_0 M_0\sin\vartheta\frac{\partial f}{\partial x_1}+C(x_1,x_2,t).$$

The unknown function $C(x_1,x_2,t)$ is determined from the boundary condition (8) for the pressure disturbance on the free surface

$$C(x_1,x_2,t)=\Sigma(x_1,x_2,t)+\eta\left.\frac{\partial v_3}{\partial x_3}\right|_{x_3=0},$$

$$\Sigma(x_1,x_2,t)=\rho g\zeta-\alpha\left(\frac{\partial^2\zeta}{\partial x_1^2}+\frac{\partial^2\zeta}{\partial x_2^2}\right)+\frac{\mu_0}{2}M_0^2\sin 2\vartheta\frac{\partial\zeta}{\partial x_1}$$

$$-\mu_0 M_0\left(\left.\frac{\partial f}{\partial x_1}\right|_{x_3=0}\sin\vartheta+\gamma\frac{\partial f}{\partial x_3}\cos\vartheta\right). \tag{11}$$

As a result, we have

$$p=\Sigma+\mu_0 M_0\sin\vartheta\frac{\partial f}{\partial x_1}+\eta\left(\frac{\partial v_3}{\partial x_3}+\left.\frac{\partial v_3}{\partial x_3}\right|_{x_3=0}\right). \tag{12}$$

After plugging the solution (12) into the projection of the equation of motion (7) onto the x_1 and x_2 axes and discarding the quantities of order ε^2, we obtain

$$\eta\frac{\partial^2 v_j}{\partial x_3^2}=\frac{\partial\sigma}{\partial x_j},\quad j=1,2,\quad \sigma=\Sigma-\mu_0 M_0\cos\vartheta\frac{\partial f}{\partial x_3}. \tag{13}$$

Solutions of these equations which satisfy the conditions for the absence of tangential stresses on the free surface (8) and the conditions for the attachment of fluid particles to the wall surface (9) have the simple form

$$v_j=\frac{x_3^2-d^2}{2\eta}\frac{\partial\sigma}{\partial x_j},\quad j=1,2. \tag{14}$$

In writing (14) we omitted the quantities of order ε^2. Allowing for (14), from the continuity equation (7) we find the distribution of the vertical velocity which satisfies the condition of zero flow (9) at the wall

$$v_3=\frac{1}{6\eta}(2d^3+3d^2x_3-x_3^3)\left(\frac{\partial^2\sigma}{\partial x_1^2}+\frac{\partial^2\sigma}{\partial x_2^2}\right). \tag{15}$$

Taking into account this expression, we can easily see that the last term on the right-hand side of (12) should be omitted, since it is of order ε^2 in comparison to the first term.

Substituting the solution (15) into the kinematic condition on the free surface, i.e., the first expression in (8), with allowance for (11) and (13) we obtain

$$\frac{\partial\zeta}{\partial t}=\frac{d^3}{3\eta}\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}\right)\left\{\rho g\zeta-\left(\frac{\partial^2\zeta}{\partial x_1^2}+\frac{\partial^2\zeta}{\partial x_2^2}\right)+\frac{\mu_0}{2}M_0^2\sin 2\vartheta\frac{\partial\zeta}{\partial x_1}-\mu_0 M_0\left[\left.\frac{\partial f}{\partial x_1}\right|_{x_3=0}\times\sin\vartheta+(1+\gamma)\frac{\partial f}{\partial x_3}\cos\vartheta\right]\right\}. \quad (16)$$

Thus, the temporal evolution of an initially assigned small disturbance of the shape of an originally flat free surface of a fluid is determined by the system of equations (4) and (16) together with the conditions for conjugation of the magnetic fields on the interfaces between media with different magnetic properties (8) and (9) and the conditions at infinity (10). After solving this problem, the pressure disturbance in the fluid layer and the velocity field are calculated using (12), (14), and (15).

ANALYSIS OF INSTABILITY

A disturbance $\zeta(x_1,x_2,0)$ of the shape of a free surface appearing in the initial moment in time and decaying fairly rapidly as $x_1,x_2\to\pm\infty$ can be represented in the form of a Fourier integral. In the linear formulation, the subsequent temporal evolution of each Fourier component is independent of the evolution of the other Fourier components, so that the solution of the problem formulated above can be constructed in the form of a superposition of partial solutions of the form

$$(\zeta,f,f_j)=\exp[i(k_1x_1+k_2x_2-\omega t)][a,F(x_3),F_j(x_3)];$$
$$i=\sqrt{-1};\quad j=1,2. \quad (17)$$

The complex amplitude functions a, F, and F_j depend not only on x_3, but also on the real parameters k_1 and k_2, i.e., the components of the wave vector $\mathbf{k}=(k_1,k_2,0)$.

After substituting the expressions (17) into the simplified system of magnetostatic equations (4) with the corresponding boundary conditions (8) and (9), we obtain

$$F''=0,\quad F_j''-k^2F_j=0,\quad j=1,2,\quad k=\sqrt{k_1^2+k_2^2},$$
$$x_3=0:\ F-F_1=M_0a\cos\vartheta,$$
$$(1+\gamma)F'-F_1'=ik_1aM_0\sin\vartheta,$$
$$x_3=-d:\ F=F_2,\quad (1+\gamma)F'=F_2'.$$

In the approximation under consideration, the solution of this problem has the form

$$F=\frac{a}{2}M_0(ik_1\sin\vartheta+k\cos\vartheta)\left(\frac{x_3}{1+\gamma}+\frac{1}{k}\right),$$
$$F_1=\frac{a}{2k}M_0(ik_1\sin\vartheta-k\cos\vartheta)\exp(-kx_3),$$
$$F_2=\frac{a}{2k}M_0(ik_1\sin\vartheta+k\cos\vartheta)\exp(kx_3). \quad (18)$$

Substituting the solution consisting of (17) and (18) into Eq. (16), we obtain the dispersion relation

$$\omega=is,$$
$$s=\frac{d^3}{3\eta}\left\{-\rho g(k_1^2+k_2^2)+\frac{\mu_0}{2}[M_{03}^2(k_1^2+k_2^2)^{3/2}-M_{01}^2k_1^2(k_1^2+k_2^2)^{1/2}]-\alpha(k_1^2+k_2^2)^2\right\}, \quad (19)$$

where $M_{01}=M_0\sin\vartheta$ and $M_{03}=M_0\cos\vartheta$ are the components of the vector of the undisturbed magnetization of the fluid layer $\mathbf{M}_0$.

With consideration of (19) and the initial condition $\zeta(x_1,x_2,0)=\zeta_0(x_1,x_2)$ the disturbance of the shape of the free surface of the fluid at $t>0$ is written in the following manner:

$$\zeta(x_1,x_2,t)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A(k_1,k_2)\exp[i(k_1x_1+k_2x_2)+s(k_1,k_2)t]dk_1dk_2,$$
$$A(k_1,k_2)=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\zeta_0(x_1,x_2)\times\exp[-i(k_1x_1+k_2x_2)]dx_1dx_2. \quad (20)$$

Hence it follows that for $s(k_1,k_2)<0$ the initial disturbance damps with the passage of time, while for $s(k_1,k_2)>0$ it grows, and this growth, as can be seen from (19), is induced by the normal component $\mathbf{M}_{0n}=M_{03}\mathbf{a}_3$ of the magnetization of the fluid layer.

It also follows from (19) that in the case where the angle φ between the tangential component $\mathbf{M}_{0\tau}=M_{01}\mathbf{a}_1$ of the magnetization and the wave vector $\mathbf{k}$ differs from $\pi/2$ and $3\pi/2$, $\mathbf{M}_{0\tau}$ weakens the destabilizing effect of $\mathbf{M}_{0n}$ on the harmonic with the wave vector $\mathbf{k}$ under consideration for a fixed $\mathbf{H}_{0\tau}$ and a quasistatic increase in the vertical component of the undisturbed field $\mathbf{H}_{0n}$ from an initial zero value. Because $\mathbf{H}_{0\tau}\neq 0$, the most favorable conditions for the appearance of instability include the harmonics for which $\varphi=\pi/2$ and $\varphi=3\pi/2$. Thus, as $\mathbf{H}_{0n}$ is increased, the harmonics with $\mathbf{k}\perp\mathbf{H}_{0\tau}$ begin to grow first.

An analysis of the dispersion relation (19) shows that instability arises when M_{03} exceeds the critical value $M_c=2\sqrt[4]{\rho g\alpha/\mu_0^2}$. For comparison, we present the critical value of the magnetization of a magnetic fluid with a flat free surface which is immersed in a vertical field and occupies the entire lower half-space:[3,4] $M_{c\infty}=\sqrt{2/\mu_0}\sqrt[4]{\rho g\alpha}\sqrt{1+(\mu_r^0\mu_{rt}^0)^{-1/2}}$, where μ_r^0 and μ_{rt}^0 are defined by (3). It follows from the formulas presented that $M_c/M_{c\infty}=\sqrt{2/[1+(\mu_r^0\mu_{rt}^0)^{-1/2}]}>1$, i.e., a thin layer of a viscous magnetic fluid is more stable than a thick layer of the same fluid.

In the case of $M_{03}/M_c>1$, using the dispersion relation (19) it is not difficult to find the characteristic time for the development of instability

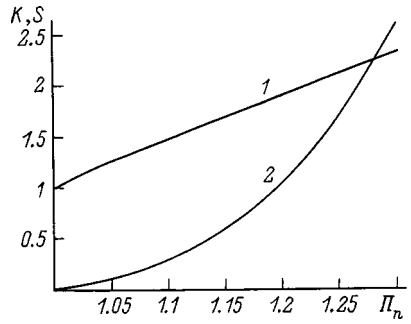

FIG. 2. Plots of $K(\Pi_n)$ (*1*) and $S(\Pi_n)$ (*2*).

$$\tau_i=\left[\max_{\substack{-\infty<k_1<\infty\\ -\infty<k_2<\infty}} \operatorname{Im}\omega(k_1,k_2)\right]^{-1}$$

and the wave vectors $\mathbf{k}_m$ of the most rapidly growing harmonics

$$\mathbf{k}_m=\pm k_m\mathbf{a}_2,\quad k_m=\sqrt{\frac{\rho g}{\alpha}}K(\Pi_n),\quad \tau_i=s^{-1}(0,\pm k_m),$$

$$s(0,\pm k_m)=\frac{\rho^2g^2d^3}{\alpha\eta}S(\Pi_n),\quad \Pi_n=\frac{M_{03}}{M_c},$$

$$K(\Pi_n)=\frac{3}{4}\Pi_n^2+\sqrt{\frac{9}{16}\Pi_n^4-\frac{1}{2}},$$

$$S(\Pi_n)=\frac{1}{6}K^2(\Pi_n)[\Pi_n^2K(\Pi_n)-1]. \tag{21}$$

Plots of the functions $K(\Pi_n)$ and $S(\Pi_n)$ are presented in Fig. 2.

With consideration of (21) the condition for applicability of the simplified equation of motion (7) is written in the following manner:

$$TS(\Pi_n)\ll 1,\quad T=\frac{\rho^3g^2d^5}{\alpha\eta^2}. \tag{22}$$

As an example of commercial magnetic fluids, let us consider a perfluorocarbon fluid, for which we have[3] $\rho=2050\ \mathrm{kg/m^3}$, $\eta=2.5\ \mathrm{N\cdot s/m^2}$, and $\alpha=0.018\ \mathrm{N/m}$. On the basis of these values we find $M_c=7.78\ \mathrm{kA/m}$, which is smaller than the table value of the saturation magnetization $M_s=7.96\ \mathrm{kA/m}$ for a fluid of this type. Thus, in the case under consideration $\Pi_n\leqslant 1.023$, and $S(1.023)=0.034$. Because this quantity is small, the layer thickness at which the constraint (22) is satisfied, can be found to an order of magnitude from the condition $T\sim 1$. As a result, we obtain $d\sim 1$ mm. When $d=1$ mm, we have $\tau_d=0.007$ s, and in the case of $M_{03}=M_s$, formulas (21) give $k_m=12\ \mathrm{cm^{-1}}$ and $\tau_i=2.7$ s.

BREAKUP OF A LAYER IN A TILTED MAGNETIC FIELD

Turning to the case of $M_{03}>M_c$, let us consider the development of an initial small disturbance of the shape of the free surface of a magnetic fluid in a tilted magnetic field. Going over to the dimensionless variables

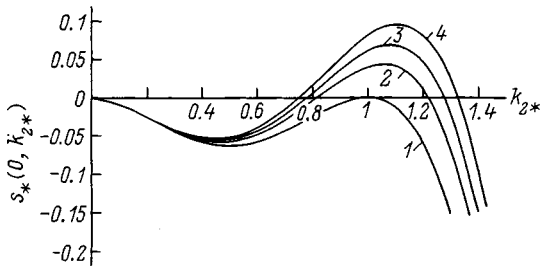

FIG. 3. Plots of $s_*(0,k_{2*})$. Π_n: *1*—1, *2*—1.01, *3*—1.015, *4*—1.02.

$$k_*=k\sqrt{\frac{\alpha}{\rho g}},\quad \omega_*=\frac{3\alpha\eta}{\rho^2g^2d^3}\omega,$$

$$s_*=\frac{3\alpha\eta}{\rho^2g^2d^3}s,\quad \Pi_\tau=\frac{M_{01}}{M_c}$$

in the dispersion relation (19), we obtain

$$\omega_*=is_*,\quad s_*=-(k_{1*}^2+k_{2*}^2)+2\sqrt{k_{1*}^2+k_{2*}^2}\times[\Pi_n^2(k_{1*}^2+k_{2*}^2)-\Pi_\tau^2k_{1*}^2]-(k_{1*}^2+k_{2*}^2)^2. \tag{23}$$

The families of $s_*(0,k_{2*})$ and $s_*(k_{1*},0)$ curves for various values of Π_n and Π_τ in the special cases of $\mathbf{k}_*=k_{2+}\mathbf{a}_2$ and $\mathbf{k}_*=k_{1*}\mathbf{a}_1$ are shown in Figs. 3 and 4. Each of the plots presented continues into the left half-plane symmetrically relative to the vertical axis. We note that curve *4* in Fig. 3 and curve *1* in Fig. 4 are identical. Curve *1* in Fig. 3 and curve *3* in Fig. 4, which correspond to $\Pi_\tau=\sqrt{\Pi_n^2-1}$ and $\Pi_n=1.02$, are also identical.

In the case of a vertical magnetic field ($\Pi_\tau=0$), Eq. (23) takes the form $s_*=\Omega(\sqrt{k_{1*}^2+k_{2*}^2},\Pi_n)$. For a fixed value of Π_n this dependence is represented geometrically by the surface formed when the curve in Fig. 3 corresponding to the value of Π_n chosen is rotated about the vertical axis. If $\Pi_n=\mathrm{const}$ the level lines of Ω are naturally a family of concentric circles with a center at the origin of coordinates in the (k_{1*},k_{2*}) plane of variables. In this case the temporal evolution of each of the harmonics appearing in the integrand in (20) is determined only by $|\mathbf{k}|$, i.e., does not depend on the direction of $\mathbf{k}$. The experiment shows[4] that the breakup of a sufficiently thin layer of a magnetic fluid immersed in a vertical magnetic field results in the formation of a system of hexagonal cells consisting of a central conical drop and six similar drops surrounding it. We note that a

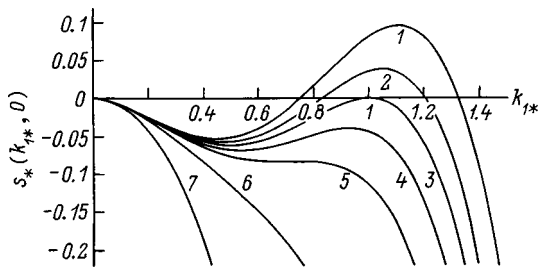

FIG. 4. Plots of $s_*(k_{1*},0)$ for the fixed value $\Pi_n=1.02$. Π_τ: *1*—0, *2*—0.15, *3*—0.201, *4*—0.25, *5*—0.312, *6*—0.5, *7*—1.02.

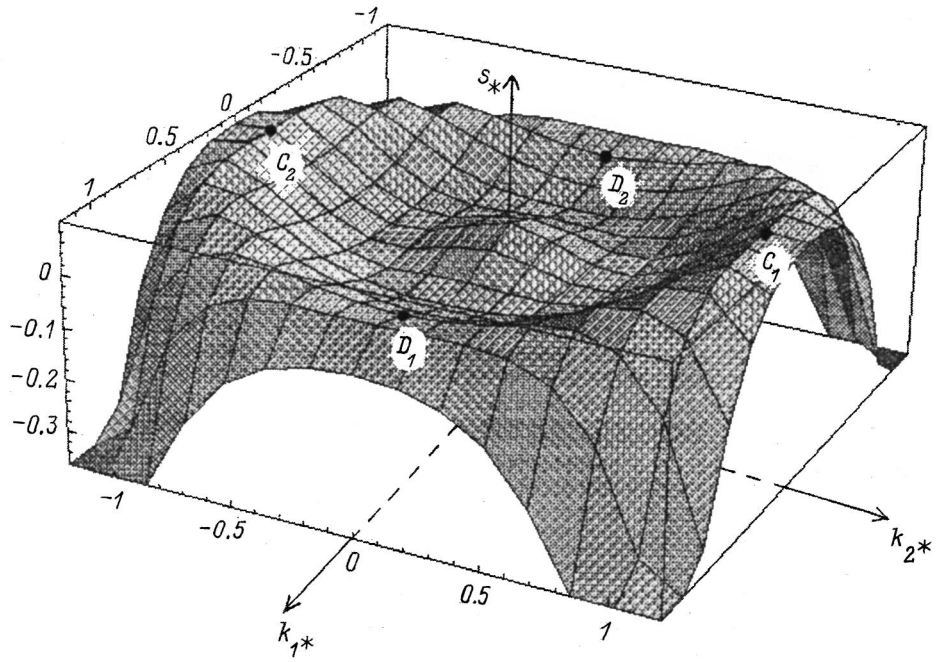

FIG. 5. Plot of the $s_*=s_*(k_{1*},k_{2*})$ surface for $\Pi_n=1.02$ and $\Pi_\tau=0.201$.

similar picture is observed when a thin layer of a viscous fluid breaks up as a consequence of Rayleigh–Taylor instability.[7]

As an example, Fig. 5 shows the $s_*=s_*(k_{1*},k_{2*})$ surface for $\Pi_n=1.02$ and $\Pi_\tau=\sqrt{\Pi_n^2-1}=0.201$, and Fig. 6 shows the level lines of $s_*(k_{1*},k_{2*})$ corresponding to the case of instability ($s_*>0$). It is seen from Figs. 5 and 6 that in the case of $\mathbf{H}_\tau\neq 0$, along with the peaks (the points $C_{1,2}=[0,\pm K(\Pi_n),S(\Pi_n)]$, which correspond to the maxima on curve *4* in Fig. 3 continued into the left half-plane) there are also saddle points at $D_{1,2}=[\pm K(\sqrt{\Pi_n^2-\Pi_\tau^2}),0,S(\sqrt{\Pi_n^2-\Pi_\tau^2})]$. If $\Pi_\tau=\sqrt{\Pi_n^2-1}$ and $\Pi_n\geqslant 1$, the saddle points correspond to the points $(\pm 1,0)$ in Fig. 4, at which curve *3* is tangent to the horizontal axis. At a fixed value of Π_n, as Π_τ is increased from zero, the saddle points on the $s_*=s_*(k_{1*},k_{2*})$ surface that correspond to the most distant critical points from the origin of coordinates on the curves in Fig. 4 pass from the upper half-space into the lower. When $\Pi_\tau=\sqrt{\Pi_n^2-2\sqrt{2}/3}$ (curve *5* in Fig. 4), because of the merging of the two critical points not at the origin of coordinates (among the three critical points on each curve in Fig. 4 when $\Pi_\tau<\sqrt{\Pi_n^2-2\sqrt{2}/3}$), the saddle points on the $s_*=s_*(k_{1*},k_{2*})$ surface cease to exist.

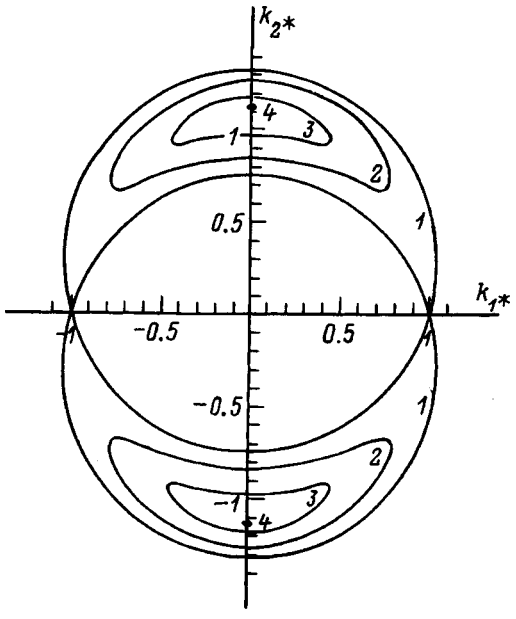

FIG. 6. Level lines of $s_*(k_{1*},k_{2*})$ for $\Pi_n=1.02$ and $\Pi_\tau=0.201$. s_*: *1*—0, *2*—0.035, *3*—0.07, *4*—0.096.

Figure 7, which shows a family of $s_*(k_{1*},k_{2*})=0$ level lines for the fixed value $\Pi_n=1.02$, illustrates the influence of the tangential component of the magnetization of the layer on the form of the instability region in the (k_{1*},k_{2*}) plane of variables. In the case of a vertical ($\Pi_\tau=0$) magnetic field exceeding the critical value, the instability region $\Pi_n^2-\sqrt{\Pi_n^4-1}<k_{1*}^2+k_{2*}^2<\Pi_n^2+\sqrt{\Pi_n^4-1}$ takes the form of an annular ring bounded by two concentric circles (curves *1*). As Π_τ is increased, the instability region, which remains connected for Π_τ in the range $0\leqslant\Pi_\tau\leqslant\sqrt{\Pi_n^2-1}$, deforms in such a way that the distance between its boundaries decreases for all $k_{1*}\neq 0$. When $\Pi_\tau=\sqrt{\Pi_n^2-1}$, the outer and inner boundaries of the original ring merge at the points which are projections of the saddle points $D_{1,2}$ onto the (k_{1*},k_{2*}) plane. As a result of such merging, the instability region takes the form of two crescents bounded by curves *3*. Just this case corresponds to the $s_*=s_*(k_{1*},k_{2*})$ surface plotted in Figure 5, as well as curve *3* in Fig. 4. When Π_τ is increased further, the instability region breaks up into two individually connected components, each of which is bounded by curves *4*–*7*.

Figure 8 presents the $s_*=s_*(k_{1*},k_{2*})$ surface for $\Pi_n=1.02$ and $\Pi_\tau=\sqrt{\Pi_n^2-2\sqrt{2}/3}=0.312$. Curves *5* in Figs. 4 and 7 correspond to the same values of Π_n and Π_τ. It can be seen from Figs. 8 and 7 that the $s_*=s_*(k_{1*},k_{2*})$ surface

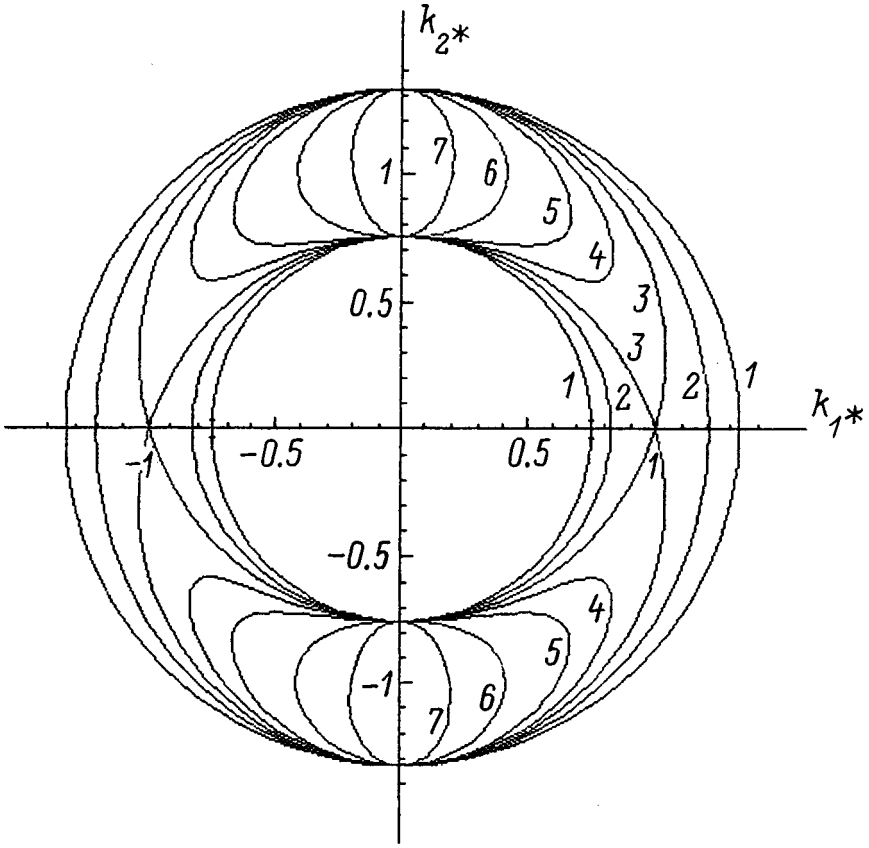

FIG. 7. Family of $s_*(k_{1*},k_{2*})=0$ level lines for the fixed value $\Pi_n=1.02$. Π_τ: *1*—0, *2*—0.15, *3*—0.201, *4*—0.25, *5*—0.312, *6*—0.5, *7*—1.02.

for $\Pi_n>1$ and $\Pi_\tau \geqslant \sqrt{\Pi_n^2-2\sqrt{2}/3}$ has two clearly expressed peaks. Consequently, at fairly large t the main contribution to the integral representation of the shape (20) of the free surface is made by the harmonics having wave vectors whose tips lie in the vicinity of the points which are the projections of $C_{1,2}$ onto the k_{1*},k_{2*} plane, i.e., the harmonics having wave vectors which form angles close to $\pi/2$ and $3\pi/2$ with $\mathbf{H}_\tau$. For $\Pi_n>1$ and $\Pi_\tau \gtrsim \sqrt{\Pi_n^2-2\sqrt{2}/3}$ this means that in the linear stage of the development of instability the shape of the free surface of the fluid varies considerably more rapidly in the direction transverse to $\mathbf{H}_\tau$ than along $\mathbf{H}_\tau$.

Thus, for $M_{03}>M_c$ and $M_{01} \gtrsim \sqrt{M_{03}^2-2\sqrt{2}M_c^2/3}$ the linear theory of the breakup of a thin layer of a viscous mag-

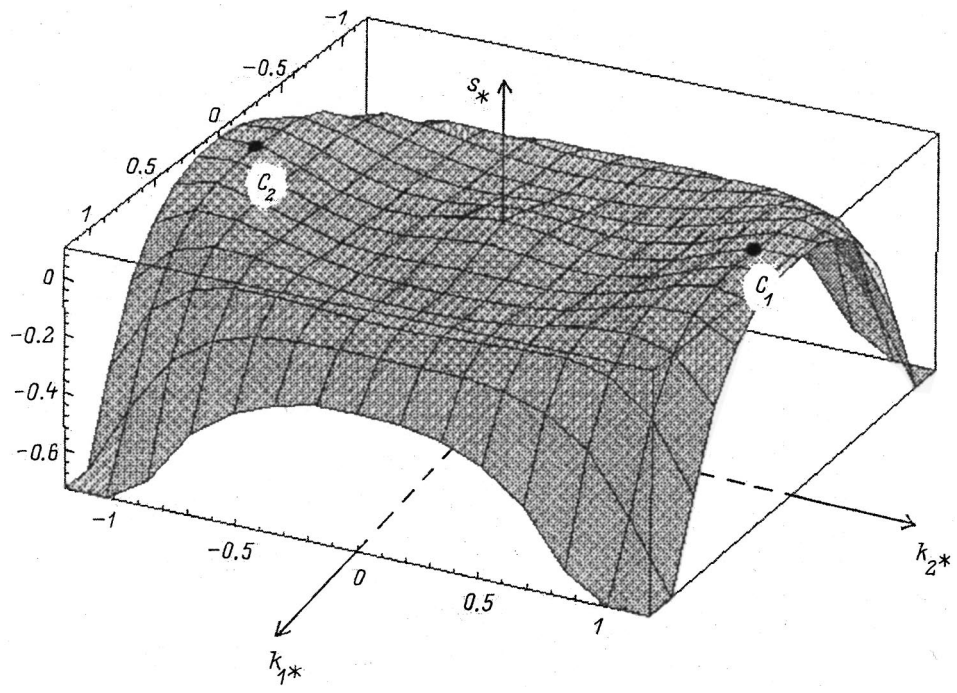

FIG. 8. Plot of the $s_*=s_*(k_{1*},k_{2*})$ surface for $\Pi_n=1.02$ and $\Pi_\tau=0.312$.

netic fluid in a tilted magnetic field predicts the formation of a system of shafts of fluid extended along $\mathbf{H}_\tau$, whose crests are separated by a distance of the order of $2\pi/k_m$. Unlike the corrugations on the interface between thick layers of magnetic fluids in a tilted field,[4] in the case under consideration the final breakup pattern of the layer will consist, roughly speaking, of a system of shafts parallel to $\mathbf{H}_\tau$, in which the magnetic fluid is localized and which are separated by stripes representing portions of the plate surface not covered by the fluid.

INSTABILITY OF A LAYER OF A FERROFLUID COVERING A MAGNETIZED PLATE IN THE PRESENCE OF AN ORTHOGONAL FIELD

Let us explore the effect of magnetizing a thick (in comparison to the characteristic wavelengths of the unstable harmonics) plate on the instability of a flat thin layer of a viscous magnetic fluid immersed in an orthogonal magnetic field. We assume that the material of the plate obeys a linear magnetization law, while the fluid is magnetized nonlinearly. In the configuration under consideration with a uniform field $\mathbf{H}_{01}=(0,0,H_{01})$ far above the fluid, the field in the fluid layer with a flat free surface $\mathbf{H}_0=(0,0,H_0)$ and the field within the plate $\mathbf{H}_{02}=(0,0,H_{02})$ are calculated from the condition of continuity of the normal component of the induction at the interfaces between media with different magnetic properties.

The potentials of the disturbances of the magnetic fields caused by deformation of the free surface are described by the system of equations (4), for which the boundary conditions have the form

$$x_3=0:\quad f-f_1=M_0\zeta,\quad \mu_{rt}^0\frac{\partial f}{\partial x_3}=\frac{\partial f_1}{\partial x_3},$$

$$x_3=-d:\quad f=f_2,\quad \mu_{rt}^0\frac{\partial f}{\partial x_3}=\mu_{r2}\frac{\partial f_2}{\partial x_3}, \tag{24}$$

where $\mu_{r2}=\mu_2/\mu_0$ is the relative permeability of the plate material.

Taking into account the representations (17), we can easily write out a solution of the problem consisting of (4) and (24) that vanishes as $x_3\to\pm\infty$. In the approximation under consideration we have

$$f=\frac{\zeta}{\mu_{r2}+1}M_0\left[\frac{\mu_{r2}}{\mu_{rt}^0}k(x_3+d)+1\right],$$

$$f_1=-\frac{\mu_{r2}\zeta}{\mu_{r2}+1}M_0\exp(-kx_3),$$

$$f_2=\frac{\zeta}{\mu_{r2}+1}M_0\exp(kx_3). \tag{25}$$

It is easily seen from the expressions (17), (18), and (25) for the same vertical field $\mathbf{H}_0=(0,0,H_0)$ and the deformation of flat free surfaces of identical layers that the disturbance of the field in the fluid covering a magnetized plate ($\mu_{r2}>1$) is greater than in the case of an unmagnetized plate ($\mu_{r2}=1$). We note that for the flat configurations with a fixed $\mathbf{H}_{01}=(0,0,H_{01})$ compared the fields $\mathbf{H}_0=(0,0,H_0)$ within the fluid layers are identical.

When the first of the expressions in (25) is substituted into Eq. (16) in the case of $\vartheta=0$ and $\gamma=\chi_t^0$, with consideration of (17) we obtain the dispersion relation

$$\omega=is,$$

$$s=\frac{d^3}{3\eta}\left[-\rho g(k_1^2+k_2^2)+\frac{\mu_0\mu_{r2}}{\mu_{r2}+1}M_0^2(k_1^2+k_2^2)^{3/2}-\alpha(k_1^2+k_2^2)^2\right]. \tag{26}$$

Using (26) to calculate the critical value M_{cm} of the magnetization of the fluid layer, we obtain

$$M_{cm}=\sqrt[4]{\rho g\alpha}\sqrt{\frac{2(\mu_{r2}+1)}{\mu_0\mu_{r2}}}.$$

In this case the wave number of the most rapidly growing harmonics and the characteristic time for the development of instability are specified, as before, by formulas (21), in which Π_n must be replaced by $\Pi_m=M_0/M_{cm}$.

Comparing the values of M_{cm} and M_c related to the cases of $\mu_{r2}>1$ and $\mu_{r2}=1$, we have $M_{cm}/M_c=\sqrt{(\mu_{r2}+1)/(2\mu_{r2})}<1$. Thus, in the case of a plate material with a high permeability ($\mu_{r2}\gg1$) the threshold for the appearance of the instability and breakup of a thin layer of a viscous magnetic fluid is roughly 30% lower than in the case of a nonmagnetizable plate.

CONCLUSION

Order-of-magnitude estimates have been made in a linearized system of ferrohydrodynamic and magnetostatic equations with boundary conditions describing three-dimensional wave motions in a thin layer of a nonlinearly magnetized viscous fluid on a horizontal nonmagnetic plate in the presence of a uniform tilted magnetic field which is fixed far from the fluid layer. The condition for realization of a fluid motion regime in which the diffusion of vorticity across the layer plays a significant role has been written out.

An equation in partial derivatives relating the disturbance of the shape of the initially flat, free surface of a magnetic fluid to the disturbance of the magnetic-field potential caused by bending of the free surface, which is an interface between media with different magnetic properties, has been derived in reference to such a case. A dispersion relation has been obtained on the basis of this equation within a formulation of the problem of the instability of a flat layer differing from the formulation in the approximation of plane-parallel potential motion of an inviscid magnetic fluid,[3,4] and this relation has been analyzed.

The critical value of the vertical component of the magnetization of the layer has been calculated, and it has been established that a thin layer of a viscous magnetic fluid is

more stable than a thick layer of the same fluid. The characteristic time for the development of instability has been found. An estimate of the value of the tangential component of the magnetization of the layer has been obtained in the case where the anisotropy of the development of unstable harmonics of the initial disturbance of the shape of the free surface due to this component causes the breakup of the continuous layer of the magnetic fluid to be accompanied by the formation of a system of shafts parallel to the tangential component of the applied field. An estimate of the characteristic distance between the shafts has been given. It has been established that magnetizing a thick plate of a material with a high permeability lowers the critical value of the vertical component of the magnetization of the fluid layer by roughly 30% (in comparison to the case of a plate composed of a nonmagnetic material).

This work was performed with support from the Russian Fund for Fundamental Research (Project No. 96-01-01747).

[1] M. D. Cowley and R. E. Rosensweig, J. Fluid Mech. **30**, 671 (1967).
[2] R. E. Zelazo and J. R. Melcher, J. Fluid Mech. **39**, 1 (1969).
[3] R. E. Rosensweig, *Ferrohydrodynamics* [Cambridge University Press, Cambridge–New York (1985); Mir, Moscow (1989), 357 pp.].
[4] B. M. Berkovsky, V. F. Medvedev, and M. S. Krakov, *Magnetic Fluids: Engineering Applications* [Oxford University Press, Oxford–New York (1993); Khimiya, Moscow (1989), 238 pp.].
[5] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. [Pergamon Press, Oxford (1987); Nauka, Moscow (1988), 736 pp.].
[6] P. H. LeBlond and F. Mainardi, Acta Mech. **68**, 203 (1987).
[7] M. Fermigier, L. Limat, and J. E. Wesfreid *et al.*, J. Fluid Mech. **236**, 349 (1992).

Translated by P. Shelnitz

Nonadiabatic excitation of iodine molecules in the translational disequilibrium zone of a shock wave

V. Yu. Velikodnyĭ, A. V. Emel'yanov, and A. V. Eremin

Scientific-Research Center for the Thermal Physics of Pulsed Effects, Joint Institute of High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia
(Submitted March 9, 1998)
Zh. Tekh. Fiz. **69**, 23–33 (October 1999)

Short-lived peaks of nonequilibrium emission are detected at 320–350 nm in shock-wave fronts in He, Ne, Ar, and H_2 containing from 0.1 to 3% iodine molecules. The effect is observed in the range of Mach numbers from 3.2 to 6.3 for initial pressures of the mixtures ranging from 133 to 2660 Pa. The emission observed is assigned to the electronic $I_2(D^3\Sigma \rightarrow B^3\Pi)$ band, which is located at excitation energies $5.45 \rightarrow 1.8$ eV, i.e., significantly above the dissociation threshold of iodine molecules (1.54 eV). An analysis of the results shows that the leading role in the excitation of iodine molecules is played by high-energy collisions in the translational disequilibrium zone of the shock wave. The best description of the experimental data is achieved for the value of the effective collision energy in the front calculated on the basis of a numerical solution of the Boltzmann equation by a modified Tamm–Mott-Smith method. The absolute values of this energy under the conditions of the experiments performed are roughly 10 times greater than the mean collision energy in the equilibrium zone behind the shock wave. The probability of nonadiabatic supercollisions of the type $I_2+I_2 \rightarrow I_2(D^3\Sigma)+I_2-6.4$eV exceeds the adiabatic values by a factor of $10^{15}-10^{20}$. [S1063-7842(99)00410-9]

INTRODUCTION

The possibility that the translational disequilibrium in a shock-wave front influences the kinetics of physicochemical processes was first put forward in Ref. 1. The essence of the phenomenon is as follows. As we know, a plane shock wave moving through a gas with a supersonic velocity V and a density jump from ρ_1 to ρ_2 sets gas into motion with slightly smaller velocities $v=V(1-\rho_1/\rho_2)$, which significantly exceed the mean thermal velocities of the molecules in the shock-wave front. As a result, the velocity distribution function of the molecules transforms from a motionless narrow ''cold'' distribution to a moving ''smeared'' high-temperature distribution, and the region where the distribution function undergoes this transformation can be regarded as a zone where a dense beam of hot molecules interacts with a less dense, cold gas medium. Under such an approach it is qualitatively clear that the mean collision energy in the frontal zone will be determined to a large extent by the velocity of the front, rather than by the thermal velocities of the molecules c_1 and c_2, and under certain conditions can significantly exceed the mean collision energies of the molecules in the equilibrium zone behind the shock wave.

Of course, real collision energies can be determined only on the basis of a detailed treatment of the transformation of the distribution function in the shock-wave front. Such an analysis was first performed in Ref. 2, where the Boltzmann equation was solved in the case of a distribution function assigned in the form of a superposition of the initial and final distributions with weighting factors that vary along the coordinate. This problem has subsequently been solved many times both by the Monte Carlo method[3,4] and on the basis of a solution of the generalized Boltzmann equation by a modified Tamm–Mott-Smith method.[5] The results of all these studies give qualitatively similar results for the mean flow characteristics.

However, the investigation of high-threshold physicochemical processes with $E \gg kT_1$ requires knowledge of the distribution functions in the high-energy region, thereby complicating the problem. Therefore, the theoretical results in this area appeared later. The possibilities of using the techniques of direct Monte Carlo numerical simulation for this purpose are fairly limited. Calculations have been performed for two-component mixtures with thresholds $E \leqslant 30kT_1$ (Refs. 6 and 7). In this case the effect is small. The factor by which the partial contribution of collisions with such energies exceeds the equilibrium contribution behind the shock wave does not surpass 10. At the same time, a numerical solution of the Boltzmann equation by the modified Mott-Smith method[8] revealed that the contribution of collisions with energies $E \cong 100-250kT_1$ can exceed the equilibrium value by a factor of 10^6-10^7.

The phenomenon just described should be manifested most strongly in the case of the propagation of shock waves in a light gas containing a small admixture of heavy molecules. In this case the light gas, in which the speed of sound is high, serves as a sort of accelerator of the slowly moving heavy molecules, and their relative velocities following collisions with molecules before the front are so high that the kinetic energies of the colliding particles can exceed several electron volts even in weak shock waves. Let us illustrate this by a simple example. If shock waves with a Mach num-

ber $M=2$ propagate in helium ($c_{\mathrm{He}}=1050$ m/s) containing a small admixture of iodine molecules ($c_{\mathrm{I}_2}=100$ m/s), the ''partial'' Mach number for iodine will be $M(\mathrm{I}_2)\approx 20$, which corresponds to mean collision energies behind the shock-wave front of the order of 1.5 eV. In fact, the collisional kinetic energy of iodine molecules colliding with a relative velocity $v\approx 1200$ m/s is of the order of 3.5 eV, so that their partial ''longitudinal temperature'' can be still greater. (Of course, such a treatment is applicable only at distances of the order of several mean free paths, i.e., until the partial distribution function of the iodine molecules goes over to the equilibrium distribution function with the helium atoms.)

It is important, however, to stress that from the practical standpoint the occurrence of such high-energy collisions is of interest only in so far as this energy can be manifested in inelastic processes, specifically in the excitation of internal degrees of freedom of the molecules. In this respect the most significant role can be played by the initiation of processes having an activation energy E significantly exceeding the equilibrium values of kT_2 behind the shock-wave front. In order to estimate the contribution of nonequilibrium collisions to the inelastic processes, we turn to the most general expression for the effective rate constant of such a process

$$k_{\mathrm{eff}}=\int_{v_{ij(\min)}}^{\infty}\int P_{\mathrm{eff}}\sigma_{\mathrm{el}}|v_{ij}|f(v_i)f(v_j)dv_i dv_j, \tag{1}$$

where σ is the elastic collision cross section, P_{eff} is the effective probability of a transition with the energy transfer E, v_{ij} is the relative velocity of the colliding particles, $v_{ij(\min)}=\sqrt{2E/\mu}$ (μ is the reduced mass), and $f(v_i)f(v_j)$ is their distribution function.

In the equilibrium zone behind the shock wave Eq. (1) takes on the classical Arrhenius form

$$k_{\mathrm{eff}}=P_{\mathrm{eff}}\sigma_{\mathrm{el}}\langle v\rangle\exp(-E/kT_2). \tag{2}$$

It is fairly clear at this point that, generally speaking, the transition probability P_{eff} should be a function not only of the collision velocity, but also of the specific energy-transfer mechanism. In particular, in the zone behind the shock-wave front, where the characteristic time of the inelastic process is much greater than the time between elastic collisions, the process can have a multiple-step character with the transfer of energy in small portions equal to ΔE and

$$P_{\mathrm{eff}}=\sum_{\Delta E=0}^{\Delta E=E}P(\Delta E), \tag{3}$$

where for single high-energy collisions in the front $P_{\mathrm{eff}}=P(E)$. It should, however, be stressed that in the studies devoted to the role of translational disequilibrium in a shock-wave front in the initiation of physicochemical conversions,[8,9] the question of the real efficiency of the transfer of such large portions of energy ($\Delta E\geqslant 1$ eV) to internal degrees of freedom in a single collision was not considered in detail, and the only models used were the very simple model of reacting hard spheres, under the assumption that for collision energies $\geqslant E$ the transition probability $P=1$ (the ''strong-collision'' model), and the adiabatic-transition model.

In fact, according to the current theories, the main mechanism for the activation of high-threshold processes is the ''weak-collision'' mechanism, under which energy is transferred to internal degrees of freedom in small portions equal to $\Delta E\ll E$, which, as a rule, do not exceed the energy of a vibrational quantum.[10,11] Obviously, such a stepwise mechanism, which requires a large number of exciting collisions, is not feasible under the conditions of the problem at hand, and the real probability of the process is determined by the partial probability of the transfer of an anomalously large portion of energy in a single collision. In the adiabatic approximation, where the transition probability decreases exponentially with increasing values of the Massey parameter $\xi=Ea/h v$ (a is the impact parameter),[12] the contribution of such processes is negligibly small even if $E=1$ eV and $v=5000$ m/s; therefore, the effective cross section for an inelastic collision with the transfer of several electron volts is determined completely by the probability of ''strong'' nonadiabatic transitions.

The most highly developed model of nonadiabatic transitions is the Landau–Zener approximation, which is based on a treatment of the potential surfaces of the quasimolecule formed from the colliding particles, but quantitative calculations of transition probabilities even for very simple atom/diatomic-molecule systems are very complicated and unreliable.[12] In the last few years the discussion of the role of ''strong'' collisions has heated up following the discovery of ''supercollisions,'' which, despite the small transition probabilities, play a significant role in the relaxation kinetics of polyatomic molecules.[13] However, the mechanism of such ''supercollisions'' has scarcely been studied both because of the difficulties in analyzing low-probability processes in trajectory calculations, on which all the theoretical treatments are based, and because of the extremely restricted amount of experimental material available.[14,15]

For this reason, the proposed experimental approach should arouse special interest, since it opens up new possibilities for investigating very different types of inelastic processes with the transfer of anomalously large portions of energy. In particular, such processes with energies equal to $1-10$ eV can include the excitation of high-lying vibrational and electronic levels, dissociation, and ionization.

If we attempt to classify the various types of effects of practical interest which might be caused by nonequilibrium collisions in a shock-wave front, we can distinguish between: 1) first-stage pulsed effects of the type

$$A+A(M)\rightarrow A^*+A(M), \tag{4}$$

which briefly populate highly excited states A^* of the original molecules A and can thus lead to the appearance of intense emission peaks in the visible and UV regions of the spectrum; 2) second-stage macroscopic effects of the type

$$A^*\rightarrow B+C, \tag{5}$$

which are manifested by the appearance of small concentrations of dissociation and/or ionization products of the origi-

nal molecules, which recombine over the course of a fairly long time; and, finally, 3) third-stage macroscopic effects of the type

$$B+A\rightarrow D^*, \quad C+A\rightarrow E+Q,$$
$$D^*+A\rightarrow B+Q, \quad \ldots\text{etc.} \tag{6}$$

which reflect the possible initiation of avalanchelike chain reactions in exothermic systems (Q is the heat of reaction), such as ignition, detonation, and electric discharge, by active particles appearing in the front.

It should be noted that the only experimental evidence of the existence of nonequilibrium effects in the translational relaxation zone of shock waves has hitherto come from observations of peaks of nonequilibrium emission from electronically excited molecules[7,16] and that the interpretations of the possible mechanisms for populating emitting states and their relationship to the evolution of the distribution function have been developed to a clearly insufficient extent in the published studies. Therefore, setting up an experimental investigation aimed at observing and thoroughly analyzing all the aforementioned types of nonequilibrium effects in a shock-wave front would be extremely timely and interesting.

In particular, the purposes of the investigation reported herein include the experimental observation of first-stage pulsed effects in a shock-wave front, an analysis of the possible energy-transfer mechanisms, and an attempt to extract quantitative data on the probabilities of ''supercollisions'' in high-energy collisions of heavy molecules.

Shock waves in inert gases (He, Ne, and Ar) and hydrogen containing a small admixture of iodine molecules were chosen as the specific object of investigation. The short-lived peaks of nonequilibrium emission in the $I_2(D^3\Sigma\rightarrow B^3\Pi)$ transition at 340 nm were recorded in the experiments. The upper level of the emitting $D^3\Sigma$ state has an energy of the order of 5 eV, which is significantly higher than the dissociation threshold of iodine molecules (1.54 eV); therefore, populating it by the successive excitation of vibrational levels in ''weak'' collisions is ruled out, and the only mechanism responsible for the appearance of nonequilibrium emission in the vicinity of 340 nm in the transitional disequilibrium zone was provided by nonadiabatic ''supercollisions.''

EXPERIMENT

The experiments were performed in a shock tube with a diameter of 50 mm, which was outfitted with equipment for emission and absorption spectroscopy having high temporal and spatial resolution. The shock-wave velocity was measured using piezoelectric sensors at three points to an accuracy no poorer than 1%. Before each shot, the shock tube was evacuated to a pressure of 2×10^{-1} Pa and purged by the gas under investigation. Iodine vapor was admitted by passing the gas under investigation through a vessel containing crystalline iodine heated to a certain temperature. The initial pressures of the mixtures investigated ranged from 133 to 2660 Pa, and the concentrations of the iodine vapor were varied in the range from 0.1% to 3%. The actual iodine content in the mixtures at the time of passage of the shock wave was monitored by measuring the equilibrium temperature and pressure behind the front. This was possible because the effective speed of sound in a mixture of a light gas with a small admixture of iodine and thus the Mach number of shock waves for a certain propagation velocity varied significantly in response to a very slight change in the iodine content. For example, a 0.1% increase in the iodine concentration led to a roughly 10% increase in the Mach number in a 1% I_2+99% He mixture.

The intensity of the spontaneous emission of iodine molecules was measured using a long-focus optical system, which provided a spatial resolution in the shock tube $\leqslant 1$ mm. The temporal resolution of the recording equipment did not exceed 0.1 μs. The measurements in the visible and near-UV regions of the spectrum were performed using a highly sensitive FÉU-171 photomultiplier, whose output signals were fed into an S9-8 digital oscillograph and were stored by a computer. The required spectral resolution was ensured by employing narrow-band interference filters, and absolute calibration of the measurements was performed using a TRU1300-2350 tungsten ribbon-filament lamp.

In the experiments emission signals were recorded in several spectral regions characteristic of various electronic bands of the iodine molecule, including a band at 900 ± 4.5 nm, which corresponds to the $A^3\Pi\rightarrow{}^1\Sigma_g^+$ transition, a band at 501 ± 2.5 nm, which corresponds to the $B^3\Pi\rightarrow{}^1\Sigma_g^+$ transition, and a band at $320-350$ nm, which corresponds to the high-lying $D^3\Sigma\rightarrow B^3\Pi$ transition.

The results of the measurements at 900 and 501 nm did not reveal significant emission peaks in the front, while sharp emission peaks with intensities exceeding the intensity of the equilibrium emission behind the shock-wave front by several orders of magnitude in some cases were detected at 340 nm. The characteristic rise time of the signals in the peaks was no greater than $0.2-0.5$ μs, and the decay time of the signals varied in the range from 0.5 to 2 μs. The emission peaks were recorded in argon at Mach numbers $M>4.5$, in neon at $M>4.2$, in helium at $M>3.7$, and in hydrogen at $M>3.2$. Figure 1(a) shows some characteristic examples of the emission signals in the band at 340 nm for Mach numbers of the shock waves $M=4.4-4.6$ in the inert gases containing about 1% iodine and $M=5.12$ in hydrogen with the same admixture. It is clearly seen that as the speed of sound in the carrier gas increases, the amplitude of the peak increases sharply, and the decay time of the signal increases appreciably in the inert gases. It is significant that in all cases the rise time of the signal is comparable to the translational relaxation time of iodine, while its decay time is significantly greater. This could be associated qualitatively with the stepwise relaxation of the excitation energy of states which are higher-lying than the observed energy levels of the $D^3\Sigma$ state.

Figure 1b compares the signals obtained in mixtures of iodine with neon at various Mach numbers. In this case increases in the emission intensity and increases in the decay time of the signal with increasing shock-wave velocity are observed. In the hottest regime ($M=6.19$, $T>4000$ K) a quasistationary signal amplitude is observed, which is assumed to be equal to the equilibrium emission intensity in the $I_2(D^3\Sigma\rightarrow B^3\Pi)$ band for the particular shock-wave pa-

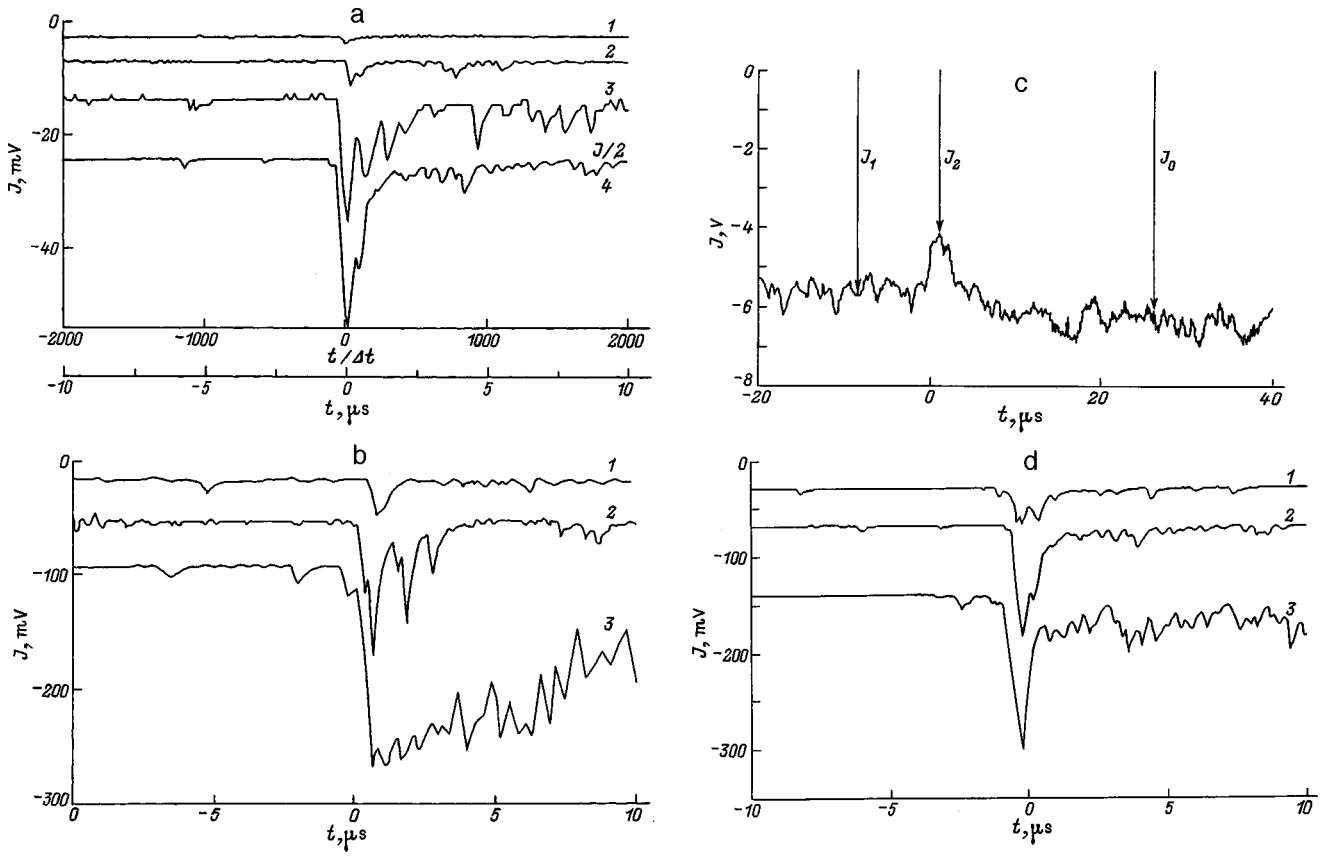

FIG. 1. a—Oscillograms of the emission signals at $\lambda = 320-340$ nm in mixtures containing about 1% I_2 in various gases for equilibrium temperatures behind the shock wave $T_2 = 2100 \pm 100$ K (Δt is the time between $I_2 + M$ collisions), M: *1*—Ar, *2*—Ne, *3*—He, *4*—H_2 (the signal in hydrogen has been reduced by a factor of 2). b—Dependence of the emission signals at $\lambda = 320-340$ nm in 1% I_2 + Ne mixtures on the Mach number of the impinging shock wave, M: *1*—4.46, *2*—5.95, *3*—6.55. c—Example of the absorption signal of iodine molecules at 187.5 ± 3.5 nm (J_0 is the total signal from the probing source, $J_0 - J_1$ is the absorption of iodine before the shock-wavefront, and $J_0 - J_2$ is the absorption of iodine behind the shock-wave front) d—Emission pulse of iodine in a shock-wave front in hydrogen: *1*—0.8% I_2 in H_2, $T_2 = 1770$ K, $P_2 = 1.73 \times 10^4$ Pa; *2*—1.2% I_2 in H_2, $T_2 = 1955$ K, $P_2 = 1.88 \times 10^4$ Pa; *3*—1.9% I_2 in H_2, $T_2 = 2507$ K, $P_2 = 2.62 \times 10^4$ Pa.

rameters. Emission oscillograms of similar form were obtained at 340 nm for all the mixtures of iodine with inert gases investigated.

In addition to the emission measurements, a special series of absorption measurements was performed in the diffuse Cordes bands of the iodine molecules in the vacuum UV region at 180–200 nm. These measurements made it possible to trace the current concentration of iodine molecules and to determine the possible influence of the thermal dissociation of I_2 behind the shock-wave front on the actual level of the emission signals. A deuterium-neon lamp, which has a continuous emission spectrum in this region, served as the radiation source, and a VM-1 monochromator served as the spectroscopic instrument. The spectral region at 187.5 ± 3.5 nm, which corresponds to the maximum absorption coefficients of the Cordes bands, was singled out in the experiments. Figure 1c presents an example of the absorption signal at 187.5 ± 3.5 nm in a 2% I_2 + 98% He mixture at $T_2 = 2750$ K and $P_2 = 4.55 \times 10^4$ Pa. The oscillogram clearly shows a sharp jump in absorption in the shock-wave front and a slow drop in absorption due to the dissociation of iodine molecules. The dissociation rate constant determined from this oscillogram is $k_d = 1.1 \times 10^{11}$ cm^3/mol·s, which agrees well with the data in the review in Ref. 17. These data on the rate of thermal dissociation of iodine show that the loss of iodine molecules in the mixtures with the inert gases during the period of $\leqslant 2$ μs following passage of the front can be neglected with a high degree of reliability over the entire temperature range investigated.

The results of the measurements in the $0.2-2\%$ $I_2 + H_2$ mixtures occupy a special place among the data obtained. In all the experiments performed at Mach numbers from 4.5 to 7, which correspond to equilibrium temperatures behind the shock-wave front in the range from 1300 to 2700 K, extremely intense emission peaks were observed in the band at 340 nm. Moreover, the decay time of the signal was no greater than 2 μs and decreased with increasing temperature [Fig. 1d]. This effect could be associated qualitatively with the rotational and vibrational relaxation of hydrogen and the effective vibronic exchange process $I_2^* + H_2 \rightarrow I_2 + H_2(v)$. Another special feature of the mixtures of iodine with hydrogen was the need to take into account the variation of the composition of the mixture as a consequence of the fast chemical reactions in the $I_2 + H_2$ system. Calculations using an expanded kinetic scheme[17] showed that at 2500 K and 5×10^4Pa the initial content of iodine molecules decreases by 10% in 5 μs and that a corresponding number of HI

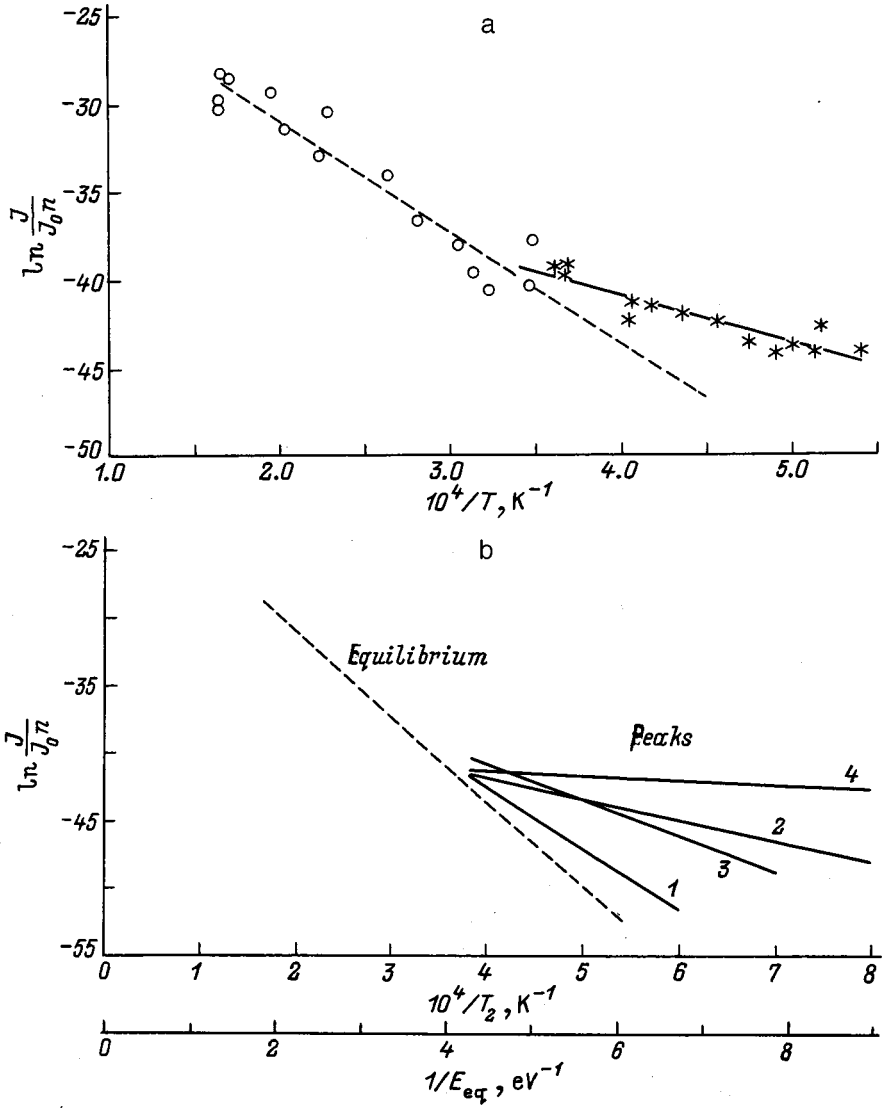

FIG. 2. Comparison of the Boltzmann dependence of the quasiequilibrium emission of iodine with an emission maximum in the peaks at 340 nm on the equilibrium temperature behind the shock wave: a—0.2–2% I_2 in helium, o-level of equilibrium emission,*—emission maximum in the peaks, b—empirical plots for various carrier gases: *1*—Ar, *2*—Ne, *3*—He, *4*—H_2.

molecules, which are optically active at 340 nm, appear during that time. The features noted, nevertheless, did not have an appreciable influence on the level of the maximum emission of the iodine molecules in the peak, which was achieved in less than 1 μs. Therefore, the analysis of the efficiency of high-energy collisions in iodine mixtures with hydrogen was distinguished from the analysis of mixtures with the inert gases only by the need to take into account the rotational relaxation of H_2, whose time can be comparable to the translational relaxation time of iodine molecules under these conditions.

DISCUSSION

The first step in the treatment of the experimental data was the construction of plots of the temperature dependence of the emission intensity in the vicinity of 340 nm. To determine the level of disequilibrium of the emission in the front, it was useful to compare the values obtained at the maxima of the peaks with the equilibrium level of the emission in the $I_2(D^3\Sigma \rightarrow B^3\Pi)$ band observed behind the most intense shock waves. The width of the spectral range singled out, which extended from 320 to 350 nm, naturally precluded resolving the vibrational structure of the band, and since the optical layer in the range singled out was known to be optically thin for all the regimes investigated, it was assumed that the equilibrium emission intensity J_{eq} in the $I_2(D^3\Sigma \rightarrow B^3\Pi)$ band is proportional to the integral population of the initial state $[I_2(D^3\Sigma)] = n^*_{eq}$, which, in turn, is related by the Boltzmann dependence

$$n^*_{eq} = n \exp(-E/kT_2) \tag{7}$$

to the total concentration of iodine molecules $[I_2] = n$ and the excitation energy E of the $I_2(D^3\Sigma)$ state. Figure 2a presents data from emission measurements in mixtures of iodine with helium, which were determined in the high-temperature region from the stationary emission level and in weaker shock waves from the emission maximum in the peaks. The measured values of J_{eq} were normalized in all cases to a constant value, viz., the emission intensity of a standard tungsten lamp J_0 in the range 320–350 nm at $T_0 = 1600$ K. The slope of the experimental dependence

$$\ln(J_{eq}/J_0 \cdot n) \propto 1/T \tag{8}$$

should reflect the effective excitation energy of the upper $I_2(D^3\Sigma)$ emitting state. The straight line drawn according to the least-squares method through the data from the equilibrium measurements of J_{eq} gave $E=5.45$ eV, which agrees fairly well with the known geometry of the iodine molecule.[18] At the same time, the data obtained at lower temperatures from the maxima of the frontal peaks J^* reveal not only a significant upward deviation of the intensity from the equilibrium value, but also a significant decrease in the slope of the temperature dependence, whose physical meaning can be associated with the relative decrease in the concentration of the kinetic energy of the shock wave to the energy of the inelastic collisions in the translational relaxation zone.

Figure 2b summarizes the experimental data on the intensities of the frontal peaks of nonequilibrium emission from iodine in various gases in the form of empirical plots of (8). Besides the scale of the reciprocal temperature T_2^{-1}, the scale of reciprocal equilibrium collision energy behind the shock wave $E_{eq}^{-1}=(kT_2)^{-1}$ is presented in the figure. It is clearly seen that an effect similar to the one noted for helium in Fig. 2a is observed in all the gases investigated, its greatest magnitude being achieved in hydrogen. It should be noted that the data obtained in the I_2+H_2 mixtures are regarded as the values of T_2 calculated under the assumption of translational-rotational equilibrium for the hydrogen molecules, although the rotational relaxation time of H_2 in such a mixture can be commensurate with the translational relaxation time of the iodine molecules. The assumption made is based on the results of a control calculation of T_2 before rotational relaxation, which led to the physically implausible result of an extremely low level of emission (3–4 orders of magnitude below the equilibrium level) even at temperatures near 4000 K. It can be theorized that the rotational relaxation rate of hydrogen on iodine molecules increased significantly as a result of vibration-rotation exchange with the iodine molecules, which have appreciable vibrational excitation already at room temperatures, according to the scheme $I_2(v)+H_2\rightarrow I_2+H_2(r)$.

In order to move on to a quantitative analysis of the nonequilibrium effects on a shock-wave front, some assumptions must be made regarding the kinetic mechanisms responsible for the appearance of the emission observed. It has already been pointed out above that the measured emission intensity in the band at 320–350 nm reflects the integral population of the $I_2(D^3\Sigma)$ state, which is determined in the general case by the balance between the excitation of iodine in single supercollisions of the type

$$I_2+I_2(\text{or}+M)\xrightarrow{k_{\text{eff}}} I_2(D^3\Sigma)+I_2(\text{or}+M) \tag{9}$$

(where M denotes the carrier gases He, Ne, Ar, and H_2) and the deactivation of excited molecules, which can occur both in quenching collisions of the type

$$I_2(D^3\Sigma)+I_2(\text{or}+M)\xrightarrow{k_q} I_2+I_2(\text{or}+M), \tag{10}$$

and as a result of the radiative transition

$$I_2(D^3\Sigma)\xrightarrow{\gamma} I_2(B^3\Pi)+h\nu. \tag{11}$$

Under the traditional kinetic approach these processes are described by the following equation for the population of excited molecules:

$$dn^*/dt=k_{\text{eff}}n^2(\text{or } k_{\text{eff}}nM)-k_qn^*n(\text{or}\,k_qn^*M)-\gamma n^*. \tag{12}$$

However, the most important feature of the processes in the translational disequilibrium zone is the fact that the effective ''rate constant'' of the excitation process (1) is an extremely strong function of time and, as will be shown below, varies significantly even during the time between elastic collisions Δt. On the other hand, the lifetime of the $D^3\Sigma$ state under investigation is $1/\gamma\approx 7$ nm (Ref. 19), and under the conditions of the experiments performed it was significantly shorter than the characteristic quenching times. Therefore, it could be assumed that the observed emission maximum is caused by the ''production'' of excited molecules in a single, very high-energy collision:

$$n^*\approx k_{\text{eff}}n^2\Delta t(\text{or } k_{\text{eff}}nM\Delta t). \tag{13}$$

In order to determine which type of exciting collisions dominates, I_2+I_2 or $I+M$, a special series of measurements of the dependence of J^* on the concentration of the carrier gas M was carried out. It was found that the emission intensity in the peaks does not depend on the concentration of the light gases H_2 and He and increases proportionally to the concentrations of the heavier carrier gases Ne and Ar. It was concluded on this basis that I_2+I_2 collisions play the leading role in exciting the iodine molecules in the lighter gases, while the I_2+Ne and I_2+Ar collisions are more effective.

The further quantitative analysis was concerned with inelastic I_2+I_2 collisions in media of H_2 and He. Under the assumption that the cross section of the nonadiabatic process (9) depends weakly on the collisional kinetic energy,[12] k_{eff} (1) can be represented in an approximation in the form

$$k_{\text{eff}}\approx P_{na}\sigma_{\text{el}}F(v_{ij}), \tag{14}$$

where σ_{el} is the elastic collision cross section, P_{na} is the probability of an inelastic transition (the nonadiabaticity factor), and

$$F(v_{ij})=\int_{v_{ij(\min)}}^{\infty}\int f(v_i)f(v_j)dv_idv_j \tag{15}$$

is the collision integral, which transforms in the case of translational equilibrium into the ''Arrhenius factor'' of the rate constant (2)

$$F_{eq}=\langle v\rangle\exp(-E/kT_2). \tag{16}$$

Figure 3 shows the evolution of the values of $F(v_{ij})$ for I_2+I_2 collisions in the translational relaxation zone of a shock wave relative to F_{eq}, which was calculated on the basis of a numerical solution of the Boltzmann equation by a modified Tamm–Mott-Smith method.[8] A comparison of the calculated $F(v_{ij})$ profiles with the experimental data in Fig. 1a reveals qualitative similarity both in the rise times of the

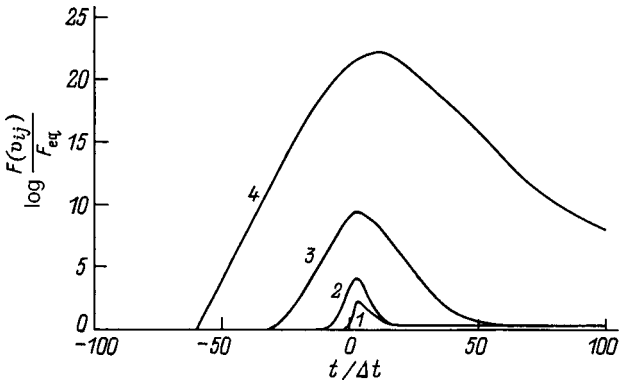

FIG. 3. Evolution of the function $F(v_{ij})/F_{\mathrm{eq}}$ [formulas (15) and (16)] for I_2+I_2 collisions in the translational relaxation zone of shock waves in mixtures of 1% I_2 with various gases for $M=4.5$ and $E=5.45$ eV. *1–4*—same as in Fig. 2b.

signals and in their dependence on the kind of carrier gas. At the same time, the characteristic decay times of the signals, as we have already noted, are roughly an order of magnitude longer than the calculated relaxation times $F(v_{ij})$ and probably characterize the integral relaxation time of the electronic energy of the iodine molecules stored in the front. Another very important conclusion following from an analysis of the plots in Fig. 3 is that $F(v_{ij})$ and thus k_{eff} vary significantly even during the time between I_2+I_2 collisions, i.e., Δt,

$$\Delta t=(n\sigma_{\mathrm{el}}\langle v\rangle)^{-1}. \tag{17}$$

Moreover, as was indicated above, the observed emission maximum is caused by the ''production'' of excited molecules in a single, very high-energy collision, whose effective energy E^* can be estimated by writing, in analogy to Ref. 16,

$$F^*(v_{ij})=\langle v\rangle\exp(-E/E^*). \tag{18}$$

Here $F^*(v_{ij})$ is the maximum value of the integral (15). These assumptions made it possible to estimate the maximum values of the concentration of excited molecules $n^*=[I_2(D^3\Sigma)]$ ''produced'' during the time Δt with neglect of the quenching (10) and the radiative processes (11)

$$n^*\approx k^*_{\mathrm{eff}}n^2\Delta t. \tag{19}$$

Taking into account (14) and (16)–(18), we obtain

$$n^*/n=P_{\mathrm{na}}\exp(-E/E^*). \tag{20}$$

Figure 4 presents the experimental values of $\ln(n^*/n)$ measured in He and H_2 as a function of the reciprocal of the effective collision energy E_i. Besides the values of $E_i=E^*$ calculated from formulas (15) and (18), the figure presents estimates of the effective collision energy $E_i=\mu v^2/2$ [$\mu=m_im_{I_2}/(m_i+m_{I_2})$ is the reduced mass] in the approximation of a molecular beam of light ($E_i=E_l\approx m_{\mathrm{He}}v^2/2, m_{H_2}v^2/2$) and heavy ($E_i=E_h=m_{I_2}v^2/4$) molecules moving with the flow velocity behind the shock wave. It is clearly seen that the use of such values of E_l and E_h leads both to significant disparity between the absolute values of n^* in He and H_2 and to obviously incorrect values of the excitation energy E of the $I_2(D^3\Sigma)$ state. At the same time, the plots of $\ln(n^*/n)\propto(E^*)^{-1}$ describe the experimental data quite well both from the standpoint of the agreement between the absolute values of n^* in different gases and with respect to the slope of the common tangent, which is closely consistent with the value obtained above $E=5.4$ eV. This result allows us to draw the fundamentally important conclusion that the model used for an approximate solution of the Boltzmann equation[8] leads to correct values of the effective energies of I_2+I_2 collisions in the wave front E^* calculated from formulas (15) and (18), and such calculations can now be used to quantitatively analyze diverse experimental data on inelastic processes in a shock-wave front.

A comparison of the values of E^* obtained with the equilibrium collision energies behind the front $E_{\mathrm{eq}}=kT_2$ [Fig. 2(b)] permits determination of the real scale of the nonequilibrium effects in the translational relaxation zone of a shock wave. In particular, it can easily be seen that the mean effective energies of the most high-energy I_2+I_2 collisions in the shock-wave fronts in He and H_2 are roughly 5–10 times greater than the equilibrium values in the flow behind the front.

The more focused view of the experimental plot of $\ln(n^*/n)\propto(E^*)^{-1}$ shown in Fig. 4b reveals that its slope increases slightly with increasing E^*. This finding can be interpreted in two ways. If we adhere to the assumption made above [see formula (13)] that the probability of the nonadiabatic transition (9) $P_{\mathrm{na}}\approx$const in the range of collision energies under consideration, the observed change in the slope can be associated with a change in the mean effective excitation energy of the iodine molecules acquired in the front. For example, under the weakest regimes with $E^*\approx0.5$ eV

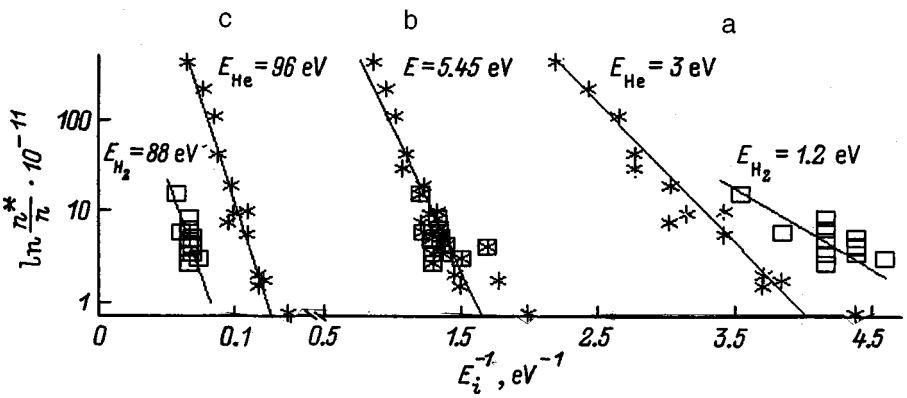

FIG. 4. Comparison of the nonequilibrium populations n^* of the $I_2(D^3\Sigma)$ state as a function of the effective collision energy E_i in shock-wave fronts in helium (*) and hydrogen (□): a—$E_i=E_l=m_{H_2}\cdot v^2/2$, $m_{\mathrm{He}}\cdot v^2/2$; b—$E_i=E^*$; c—$E_i=E_h=m_{I_2}\cdot v^2/4$.

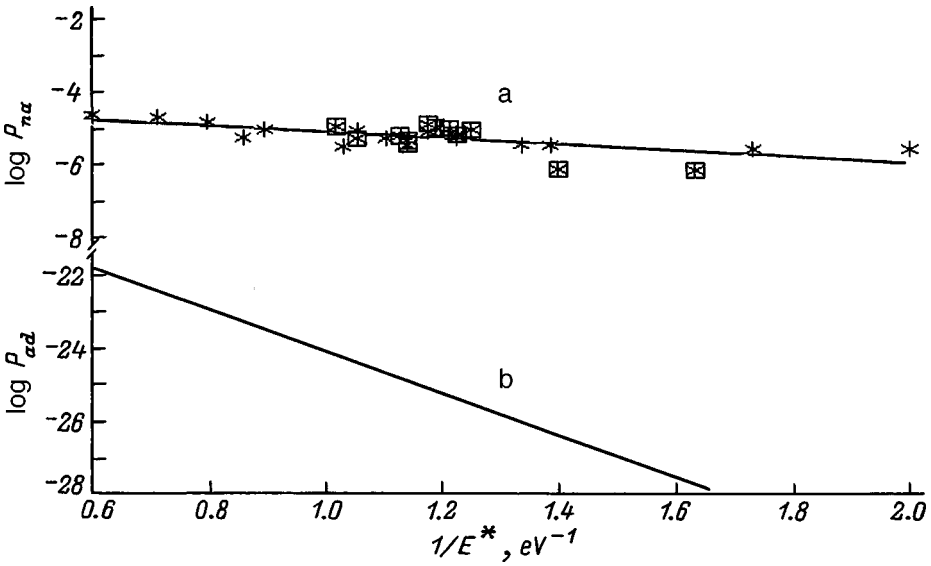

FIG. 5. Probabilities of nonadiabatic supercollisions of the type $I_2+I_2 \rightarrow I_2(D^3\Sigma)+I_2$: a—determined from experimental data using Eq. (21); b—adiabatic transition probability calculated from Eq. (22).

the observed value of E is close to the zero-point energy of the $I_2(D^3\Sigma)$ state $E' \cong 4.8$ eV, while in the most intense shock waves, where the collision energy in the front reaches $E^* \approx 1$ eV, the mean excitation energy amounts to $E'' \cong 8.5$ eV, which is close to the dissociation barrier of the $I_2(D^3\Sigma)$ state. Such a conclusion can serve as indirect confirmation of the assumption made above that the long decay time of the observed emission peaks reflects the stepwise relaxation of the higher-lying iodine states populated in the front.

On the other hand, the variation of the slope of the plots in Fig. 4 can be associated specifically with the variation of the transition probabilities as a function of the collisional kinetic energy. Under such an assumption the current values of $P_{\rm na}$ can be extracted directly from experimental measurements. In fact, after determining the ratio of the emission intensity in the peak J^* to the intensity obtained by extrapolating the corresponding equilibrium dependence $J_{\rm eq}$ [Fig. 2b] and taking into account (7) and (20), we obtain

$$\frac{J^*}{J_{\rm eq}} \cong \frac{n^*}{n^*_{\rm eq}} = \frac{P_{na}\exp(-E/E^*)}{\exp(-E/kT_2)}. \tag{21}$$

Figure 5 compares the experimental values of $P_{\rm na}$ with the effective adiabatic transition probability

$$P_{ad} \cong (1/F(v_{ij})) \times \int_{v_{ij}(\min)}^{\infty} \int \exp(-\xi)|v_{ij}|f(v_i)f(v_j)dv_i dv_j \tag{22}$$

calculated with consideration of the Massey parameter $\xi = Ea/hv_{ij}$ ($E = 5.45$ eV, $a \cong 2$ Å is the impact parameter[20]). Both probabilities depend exponentially on the collision energy and can be approximated by the simple relations

$$P_{na} = 1.75\times 10^{-4}\exp(-2.66/E^*),$$

$$P_{ad} = 4.61\times 10^{-19}\exp(-13.27/E^*).$$

It is noteworthy that besides the extremely large difference between the absolute values of $P_{\rm na}$ and $P_{\rm ad}$, which can reach 10^{29}(!), the calculated values of the adiabatic transition probability $P_{\rm ad}$ exhibits an effective energy threshold, which is more than twice as great as $E = 5.45$ eV and can be associated with the increase in the contribution of the high-energy ''tail'' of the distribution function with increasing E^*. On the other hand, the real transition probability $P_{\rm na}$ depends quite weakly on the collision energy, i.e., it has a threshold at roughly one-half of E. Qualitatively, these facts are evidence that the I_2+I_2 collisions are a certain type of nonadiabatic supercollisions, which most probably involve the formation of the intermediate complex I_4. Obviously, a quantitative analysis of the mechanisms of nonadiabatic transitions in $I_2+I_2 \rightarrow I_4+I_2(D^3\Sigma)+I_2$ collisions based on an examination of the potential surfaces of I_4 is beyond the scope of the present work and can be the subject of a special theoretical treatment.

CONCLUSION

1. Sharp emission peaks have been recorded experimentally at 320–350 nm in shock-wave fronts in inert gases and hydrogen with an admixture of 0.1–3% molecular iodine. This emission has been assigned to the $I_2(D^3\Sigma \rightarrow B^3\Pi)$ band, which is located at the excitation energies 5.45 $\rightarrow$1.8 eV, i.e., is significantly above the dissociation threshold of iodine (1.54 eV).

2. The emission intensity in the peaks increase significantly with increasing speed of sound in the carrier gas for shock waves with a fixed Mach number. The maximum emission intensities in the lightest gases (He and H_2) exceed the equilibrium values by a factor of 10^3-10^5 at Mach numbers in the range $3.2 \leqslant M \leqslant 5.5$.

3. The emission intensity of the iodine molecules did not depend on the concentration of He and H_2, but it increased proportionally to the concentrations of Ne and Ar. It was

concluded on this basis that the leading role in exciting the iodine molecules is played by the I_2+I_2 collisions in the lightest gases (He and H_2), while the I_2+Ne and I_2+Ar collisions are more effective in the heavier gases.

4. A comparison of the experimental dependences of the intensities of the emission peaks of iodine on the shock-wave velocity in He and H_2 showed that the best description of the data obtained is achieved for the value of the effective energy E^* of the I_2+I_2 collisions in the front calculated on the basis of a numerical solution of the Boltzmann equation by a modified Tamm–Mott-Smith method. The absolute values of this energy under the conditions of the experiments performed was roughly 10 times greater than the mean collision energy behind the shock wave kT_2.

5. The probability of the $I_2(^1\Sigma_g^+ \rightarrow D^3\Sigma)$ transition extracted from an analysis of the experimental data exceeded the values calculated in the adiabatic approximation by 15–20 orders of magnitude. Thus, the exciting I_2+I_2 collisions definitely have a nonadiabatic character and are most likely a certain type of complex-forming supercollisions.

As a final note, we thank T. V. Bashenov (Scientific-Research Center for the Thermal Physics of Pulsed Effects of the Russian Academy of Sciences) and A. I. Osipov (Moscow State University) for some helpful discussions of the results obtained.

This work was financially supported by the Russian Fund for Fundamental Research (RFBR), Grant No. 96-02-19694a.

[1] Ya. B. Zel'dovich, A. P. Genich, and G. B. Manelis, Dokl. Akad. Nauk SSSR, **248**, 349 (1979) [Sov. Phys. Dokl. **24**, 756 (1979)].
[2] H. Mott-Smith, Phys. Rev. **82**, 885 (1951).
[3] G. A. Bird, Phys. Fluids **13**, 1172 (1970).
[4] C. Cercignani, *Theory and Application of the Boltzmann Equation* [American Elsevier, New York (1975); Mir, Moscow (1978)].
[5] V. V. Struminskiĭ and V. Yu. Velikodnyĭ, Dokl. Akad. Nauk SSSR **266**, 28 (1982) [Sov. Phys. Dokl. **27**, 659 (1982)].
[6] G. A. Bird, *Rarefied Gas Dynamics: Proceedings of the 14th International Symposium*, Univ. of Tokyo Press (1984), Vol. 1, pp. 175–182.
[7] A. P. Genich, S. V. Kulikov, G. V. Manelis, and S. L. Chereshnev, Preprint of the Institute of Chemical Physics, Academy of Sciences of the USSR, Chernogolovka (1991).
[8] V. Yu. Velikodnyĭ and V. A. Bityurin, Pis'ma Zh. Tekh. Fiz. **22**(4), 39 (1996) [Tech. Phys. Lett. **22**(2), 150 (1996)].
[9] S. V. Kulikov, Shock Waves **7**, 25 (1997).
[10] J. Troe, *Fast Reactions in Energetic Systems*, edited by C. Cepellos and R. F. Walker, Reidel, Dordrecht–Boston (1981), pp. 125–139.
[11] A. V. Eremin, I. S. Zaslonko, and V. V. Shumova, Kinet. Katal. **37**, 485 (1996) [Kinet. Katal. **37**, 455 (1996)].
[12] E. E. Nikitin, *Theory of Elementary Atomic and Molecular Processes in Gases* [Clarendon Press, Oxford (1974); Nauka, Moscow (1970), 456 pp.].
[13] T. Lenzer, K. Luther, I. Troe *et al.*, J. Chem. Phys. **103**, 626 (1995).
[14] V. Bernshtein, I. Oref, and G. Lendvay, J. Phys. Chem. **100**, 9738 (1996).
[15] A. S. Mullin, C. A. Michaels, and G. W. Flynn, J. Phys. Chem. **102**, 6032 (1995).
[16] P. V. Kozlov, S. A. Losev, Yu. V. Romanenko, and O. P. Shatalov, Preprint No. 27-29 of the Institute of Mechanics, Moscow State University, Moscow (1997).
[17] D. L. Baulch, J. Duxbury, S. J. Grant, and D. C. Montague, J. Phys. Chem. Ref. Data Suppl. **10**, Suppl. 1, 1–1 (1981).
[18] M. McCusker, in *Excimer Lasers*, edited by C. K. Rhodes, Springer-Verlag, Berlin–Heidelberg (1979), pp. 69–117.
[19] M. Sauer, W. Mulac, R. Cooper, and F. Grieser, J. Chem. Phys. **64**, 4587 (1976).
[20] E. I. Dashevskaya, E. E. Nikitin, and I. Oref, J. Chem. Phys. **97**, 9397 (1993)

Translated by P. Shelnitz

Stability of a charged drop of a viscous electrically conducting liquid in a viscous electrically conducting medium

S. L. Shiryaeva, A. I. Grigor'ev, and D. F. Belonozhko

P. G. Demidov Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted November 3, 1997; resubmitted March 2, 1998)
Zh. Tekh. Fiz. **69**, 34–42 (October 1999)

A dispersion relation is derived for capillary oscillations of a charged electrically conducting viscous drop in an electrically conducting viscous medium. It is shown that aperiodic instability of the charged interface between the two media can arise in this system, with a growth rate that depends qualitatively differently on the ratio of their conductivities in different ranges of values of this ratio. In a certain range of conductivity ratios the drop undergoes oscillatory instability. © *1999 American Institute of Physics.* [S1063-7842(99)00510-3]

Investigation of the electrostatic instability of a charged drop of viscous liquid suspended in a different viscous liquid is of great interest in connection with many applications in which this effect appears: in technologies for the uniform mixing of immiscible liquids, in the practical problem of mixing fuel and oxidizer for the burning of liquid fuels, and in geophysical experiments (see, for example, Refs. 1–4 and the literature cited therein). Nonetheless, many questions relating to this problem have as yet been investigated very little because of the complicated experimental technique and the difficulty of the theoretical calculations. This refers also to the effect of the finite rates of charge redistribution in the two liquids on the mechanisms leading to instability of the charged surface of a drop.[5]

1. We shall consider a system consisting of two immiscible incompressible liquids with densities $\rho^{(1)}$ and $\rho^{(2)}$ and kinematic viscosities $\nu^{(1)}$ and $\nu^{(2)}$. Because of surface tension, which we denote by α, the interior liquid, which we label with an index 1, will assume the form of a spherical drop with radius R. We assume further that the liquid in the drop possesses conductivity σ_1 and permittivity ε_1, while the exterior medium is conducting with the constants σ_2 and ε_2. Let a spherical electrode of radius $R_0 \ll R$, such that the effect of the electrode on the flow of the liquid in the drop is negligibly small,[6,7] be located at the center of the drop. We assume the electrode potential to be constant, φ_0. Let a constant current, giving rise to accumulation of electric charge Q at the interface between the media, flow between the central spherical electrode and a spherical counterelectrode, located at infinity in the exterior medium and maintained at zero potential. We shall seek the spectrum of capillary oscillations of the interface.

In a spherical coordinate system (r,Θ,φ) with origin at the center of the undisturbed drop, the equation for the interface between the two liquids has the form

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

The mathematical formulation of the problem formulated, linearized around the equilibrium state, consists of the system of equations of electrodynamics[8,9]

$$\frac{\partial \mathbf{u}^{(i)}}{\partial t} = -\frac{1}{\rho^{(i)}}\nabla p^{(i)} + \nu^{(i)}\Delta \mathbf{u}^{(i)}; \quad i = 1,2; \tag{1}$$

$$\operatorname{div}\mathbf{u}^{(i)} = 0; \quad \operatorname{div}\mathbf{j}_i = 0; \tag{2}$$

$$\mathbf{j}_i = \sigma_i \mathbf{E}_i, \quad \mathbf{E}_i = -\nabla\Phi_i, \tag{3}$$

and the boundary conditions for them

$$r = R_0: \quad \Phi_1 = \varphi_0; \tag{4}$$

$$r = R + \xi: \quad u_\Theta^{(1)} = u_\Theta^{(2)}; \tag{5}$$

$$u_\varphi^{(1)} = u_\varphi^{(2)}; \tag{6}$$

$$u_r^{(1)} = u_r^{(2)} = \frac{\partial \xi}{\partial t}; \tag{7}$$

$$\Pi_{1\Theta} + \rho^{(1)}\nu^{(1)}\left[\frac{1}{r}\frac{\partial u_r^{(1)}}{\partial \Theta} + \frac{\partial u_\Theta^{(1)}}{\partial r} - \frac{1}{r}u_\Theta^{(1)}\right] = \Pi_{2\Theta} + \rho^{(2)}\nu^{(2)}\left[\frac{1}{r}\frac{\partial u_r^{(2)}}{\partial \Theta} + \frac{\partial u_\Theta^{(2)}}{\partial r} - \frac{1}{r}u_\Theta^{(2)}\right]; \tag{8}$$

$$\Pi_{1\varphi} + \rho^{(1)}\nu^{(1)}\left[\frac{1}{r\sin\Theta}\frac{\partial u_r^{(1)}}{\partial \varphi} + \frac{\partial u_\varphi^{(1)}}{\partial r} - \frac{1}{r}u_\varphi^{(1)}\right] = \Pi_{2\varphi} + \rho^{(2)}\nu^{(2)}\left[\frac{1}{r\sin\Theta}\frac{\partial u_r^{(2)}}{\partial \varphi} + \frac{\partial u_\varphi^{(2)}}{\partial r} - \frac{1}{r}u_\varphi^{(2)}\right]; \tag{9}$$

$$-p^{(1)} + 2\rho^{(1)}\nu^{(1)}\frac{\partial u_r^{(1)}}{\partial r} + p_\alpha - p_E = -p^{(2)} + 2\rho^{(2)}\nu^{(2)}\frac{\partial u_r^{(2)}}{\partial r}; \tag{10}$$

$$\Phi_1 = \Phi_2; \tag{11}$$

$$\frac{\partial \varkappa}{\partial t} + \mathbf{n}\cdot(\sigma_2\mathbf{E}_2 - \sigma_1\mathbf{E}_1) + \operatorname{div}_\Sigma(\varkappa\mathbf{u}_\tau + \varkappa b\mathbf{E}_\tau - D_s\nabla\varkappa) = 0;$$

$$\varkappa = \frac{1}{4\pi}\mathbf{n}\cdot(\varepsilon_2\mathbf{E}_2 - \varepsilon_1\mathbf{E}_1); \tag{12}$$

$$r\to\infty: \quad \Phi_2\to 0; \tag{13}$$

the condition that the volumes of the liquids are constant

$$\int_\Omega \xi(\Theta,\varphi,t)\,d\Omega = 0; \tag{14}$$

the condition that the center of mass of the system is stationary

$$\int_\Omega \xi(\Theta,\varphi,t)\mathbf{e}_r\,d\Omega = 0, \tag{15}$$

where $\xi(\Theta,\varphi,t)$, $\mathbf{u}(\mathbf{r},t)$, and $p(\mathbf{r},t)$ are the perturbations of the shape of the interface, the velocity field, and the pressure field, respectively; $\varkappa$ is the surface charge density at the interface; b and D_s are the surface mobility and the surface diffusion coefficient of charge carriers; div_Σ is the surface divergence;[8] $p_\alpha(\xi)$ is the surface-tension pressure

$$p_\alpha(\xi) = -\frac{\alpha}{R^2}(2+\Delta_\Omega)\xi; \tag{16}$$

p_E is the perturbation of the electric-field pressure P_E, related with the capillary deformation of the interface; $\Pi_{i\Theta}$ and $\Pi_{i\varphi}$ are the electric components of the tangential component of the stress tensors due to the currents of the charge redistributed as a result of deformation of the interface between the media accompanying thermal capillary motion of the liquid in both media,[9,10]

$$P_E = \frac{1}{8\pi}[\varepsilon_2 E_{2n}^2 - \varepsilon_1 E_{1n}^2] + (\varepsilon_1-\varepsilon_2)\frac{E_{2\tau}^2}{8\pi};$$

$$\Pi_{i\tau} = \frac{1}{4\pi}E_{in}E_{i\tau}; \quad i=1,2; \tag{17}$$

E_{in} and $E_{i\tau}$ are, respectively, the normal and tangential components of the electric field vector; $\mathbf{n}$ and $\tau\equiv\mathbf{e}_\Theta$, $\mathbf{e}_\varphi$ are mutually orthogonal unit vectors of the normal and two tangent vectors with respect to the surface of the drop; Δ is the Laplacian operator, and Δ_Ω is its angular part in a spherical coordinate system; and, $d\Omega$ is an element of solid angle.

We shall construct the solution in dimensionless variables, in which $R=1$, $\alpha=1$ and $\rho^{(1)}$ or $\rho^{(2)}=1$. We retain the freedom of choosing the scaling parameter, which has the dimensions of density. In the discussions below we retain the density of both media and decide which density $\rho^{(1)}$ or $\rho^{(2)}$ should be set equal to 1 after the final expressions have been obtained. All other quantities (for which we retain the old notation) will be expressed in units of their characteristic values.

2. We shall now scalarize the problem, using the procedure developed in detail in Refs. 9 and 10. The desired velocity field $\mathbf{u}^{(i)}(\mathbf{r},t)$ can be decomposed into a sum of three orthogonal fields, which we shall enumerate with the index j, assuming the values $j=1,2$, and 3,

$$\mathbf{u}^{(i)}(\mathbf{r},t) = \nabla\Psi_1(\mathbf{r},t) + (\nabla\times\mathbf{r})\Psi_2(\mathbf{r},t) + ((\mathbf{r}\times\nabla)\times\nabla)\Psi_3(\mathbf{r},t), \quad i=1,2, \tag{18}$$

where Ψ_j are scalar fields.

Substituting the expansion (18) into the incompressibility condition (2), we obtain

$$\Delta\Psi_1^{(i)} = 0. \tag{19}$$

Substituting the expansion (18) into the Navier–Stokes equation (1) gives three independent equations

$$\frac{1}{\rho^{(i)}}p^{(i)}\delta_{1j} + \frac{\partial\Psi_j^{(i)}}{\partial t} - \nu^{(i)}\Delta\Psi_j^{(i)} = 0. \tag{20}$$

Taking condition (19) into account, we obtain the scalar analog of Eqs. (1) and (2) which consists of independent equations for the scalar fields $\Psi_j^{(i)}$

$$\Delta\Psi_j^{(i)} - (1-\delta_{1j})\frac{1}{\nu^{(i)}}\frac{\partial\Psi_j^{(i)}}{\partial t} = 0 \tag{21}$$

and expressions for the pressure field $p^{(i)}$

$$p^{(i)} = -\rho^{(i)}\frac{\partial\Psi_1^{(i)}}{\partial t}. \tag{22}$$

We now substitute the expansion (18) into the boundary conditions (5) and (6), and after uncomplicated mathematical transformations we obtain

$$\left[\Psi_1^{(1)} + \frac{\partial}{\partial r}r\Psi_3^{(1)}\right] - \left[\Psi_1^{(2)} + \frac{\partial}{\partial r}r\Psi_3^{(2)}\right] = 0, \tag{23}$$

$$\Psi_2^{(1)} - \Psi_2^{(2)} = 0. \tag{24}$$

Similarly, we obtain the scalar analog of the boundary conditions (8) and (9)

$$\left[\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}(\sin\Theta[\Pi_{1\Theta}-\Pi_{2\Theta}]) + \frac{1}{\sin\Theta}\frac{\partial}{\partial\varphi}(\Pi_{1\varphi}-\Pi_{2\varphi})\right] + \rho^{(1)}\nu^{(1)}\left[2\frac{\partial}{\partial r}\frac{\Psi_1^{(1)}}{r} + \frac{\partial^2\Psi_3^{(1)}}{\partial r^2} - \frac{(2+\Delta_\Omega)}{r^2}\Psi_3^{(1)}\right] - \rho^{(2)}\nu^{(2)}\left[2\frac{\partial}{\partial r}\frac{\Psi_1^{(2)}}{r} + \frac{\partial^2\Psi_3^{(2)}}{\partial r^2} - \frac{(2+\Delta_\Omega)}{r^2}\Psi_2^{(2)}\right] = 0; \tag{25}$$

$$\left[\frac{1}{\sin\Theta}\frac{\partial}{\partial\varphi}[\Pi_{1\Theta}-\Pi_{2\Theta}] - \frac{1}{\sin\Theta}\frac{\partial}{\partial\varphi}(\sin\Theta[\Pi_{1\varphi}-\Pi_{2\varphi}])\right] + \rho^{(1)}\nu^{(1)}\left[r\frac{\partial}{\partial r}\frac{\Psi_2^{(1)}}{r}\right] - \rho^{(2)}\nu^{(2)}\left[r\frac{\partial}{\partial r}\frac{\Psi_2^{(1)}}{r}\right] = 0. \tag{26}$$

The scalar analogs of the boundary conditions (7) and (10) are obtained immediately by substituting the expansion (18) into the expression for the pressure field (22)

$$\frac{\partial\Psi_1^{(1)}}{\partial r} - \frac{1}{r}\Delta_\Omega\Psi_3^{(1)} = \frac{\partial\Psi_1^{(2)}}{\partial r} - \frac{1}{r}\Delta_\Omega\Psi_3^{(2)} = \frac{\partial\xi}{\partial t}; \tag{27}$$

$$\left[\rho^{(1)}\frac{\partial\Psi_1^{(1)}}{\partial t}+2\rho^{(1)}\nu^{(1)}\frac{\partial^2\Psi_1^{(1)}}{\partial r^2}-2\rho^{(1)}\nu^{(1)}\Delta_\Omega\frac{\partial}{\partial r}\frac{\Psi_3^{(1)}}{r}\right]$$

$$+p_\alpha-p_E=\left[\rho^{(2)}\frac{\partial\Psi_1^{(2)}}{\partial t}+2\rho^{(2)}\nu^{(2)}\frac{\partial^2\Psi_1^{(2)}}{\partial r^2}\right.$$

$$\left.-2\rho^{(2)}\nu^{(2)}\Delta_\Omega\frac{\partial}{\partial r}\frac{\Psi_3^{(2)}}{r}\right]. \tag{28}$$

The constant-volume condition (14) remains unchanged, and substituting into the condition (15), expressing that the center of mass is stationary, the expansion of the unit vector of the spherical coordinate system $\mathbf{e}_r$ in terms of unit vectors of the Cartesian system we obtain

$$\int_\Omega \xi Y_{1n}d\Omega=0. \tag{29}$$

We can see from the system obtained that the field Ψ_2 has no effect on the change in the shape of the surface, since the problem of finding it does not depend on Ψ_1, Ψ_3, and ξ and it does not appear in the equations for Ψ_1, Ψ_2, and ξ.

3. We now proceed to the solution of an ordinary scalar boundary-value problem. We seek the solutions of Eqs. (21) and the function ξ in the form of the expansions

$$\Psi_j^{(i)}=\sum_{m,n}\Psi_j^{(i)}(r)Y_{mn}\exp(St),$$

$$\xi=\sum_{m,n}ZY_{mn}\exp(St), \tag{30}$$

where S is a complex frequency, $Y_{mn}\equiv Y_{mn}(\Theta,\varphi)$ are spherical harmonics, the index m in general assumes values from 0 to ∞, and the index n assumes values from $-m$ to m.

To simplify the notation, we omit the indices n and m in the quantities $\Psi_j^{(i)}(r)$ and Z. From the conditions (14) and (29) we obtain $m\neq0$ and $m\neq1$. Substituting the expansions (30) into expression (16) for the surface-tension pressure gives

$$p_\alpha=\sum_{m,n}(m-1)(m+2)ZY_{mn}\exp(St). \tag{31}$$

We now substitute expressions (30) into Eq. (21) and find the equation for the radial functions

$$\frac{d^2\Psi_j^{(i)}(r)}{dr^2}+\frac{2}{r}\frac{d\Psi_j^{(i)}(r)}{dr}$$

$$-\left[\frac{S}{\nu^{(i)}}(1-\delta_{1j})+\frac{m(m+1)}{r^2}\right]\Psi_j^{(i)}(r)=0. \tag{32}$$

Substituting expressions (30) and (31) into Eqs. (23)–(28), we obtain the boundary conditions for the functions $\Psi_j^{(i)}(r)$

$$\left[\Psi_1^{(1)}(r)+\frac{\partial}{\partial r}r\Psi_3^{(1)}(r)\right]-\left[\Psi_1^{(2)}(r)+\frac{\partial}{\partial r}r\Psi_3^{(2)}(r)\right]=0;$$

$$\int_\Omega\left[\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}(\sin\Theta[\Pi_{1\Theta}-\Pi_{2\Theta}])+\frac{1}{\sin\Theta}\frac{\partial}{\partial\varphi}(\Pi_{1\varphi}\right.$$

$$\left.-\Pi_{2\varphi})\right]Y_{mn}d\Omega+\rho^{(1)}\nu^{(1)}\left[2\frac{\partial}{\partial r}\frac{\Psi_1^{(1)}(r)}{r}+\frac{\partial^2\Psi_3^{(1)}(r)}{\partial r^2}\right.$$

$$\left.+(m-1)(m+2)\Psi_3^{(1)}(r)\right]-\rho^{(2)}\nu^{(2)}\left[2\frac{\partial}{\partial r}\frac{\Psi_1^{(2)}(r)}{r}\right.$$

$$\left.+\frac{\partial^2\Psi_3^{(2)}(r)}{\partial r^2}+(m-1)(m+2)\Psi_3^{(2)}(r)\right]=0;$$

$$\left[\frac{\partial\Psi_1^{(1)}(r)}{\partial r}+m(m+1)\Psi_3^{(1)}(r)\right]$$

$$-\left[\frac{\partial\Psi_1^{(2)}(r)}{\partial r}+m(m+1)\Psi_3^{(2)}(r)\right]=0;$$

$$\frac{\partial\Psi_1^{(1)}(r)}{\partial r}+m(m+1)\Psi_3^{(1)}(r)-SZ=0;$$

$$\rho^{(1)}\left[S\Psi_1^{(1)}(r)+2\nu^{(1)}\frac{\partial^2\Psi_1^{(1)}(r)}{\partial r^2}\right.$$

$$\left.+2\nu^{(1)}m(m+1)\frac{\partial}{\partial r}\frac{\Psi_3^{(1)}(r)}{r}\right]-\rho^{(2)}\left[S\Psi_1^{(2)}(r)\right.$$

$$\left.+2\nu^{(2)}\frac{\partial^2\Psi_1^{(2)}(r)}{\partial r^2}+2\nu^{(2)}m(m+1)\frac{\partial}{\partial r}\frac{\Psi_3^{(2)}(r)}{r}\right]$$

$$+(m-1)(m+2)Z-\int_\Omega P_EY_{mn}d\Omega=0;$$

$$\Psi_2^{(1)}(r)-\Psi_2^{(2)}(r)=0;$$

$$\int_\Omega\left[\frac{1}{\sin\Theta}\frac{\partial}{\partial\varphi}(\Pi_{1\Theta}-\Pi_{2\Theta})-\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}(\sin\Theta[\Pi_{1\varphi}\right.$$

$$\left.-\Pi_{2\varphi}])\right]Y_{mn}d\Omega+\rho^{(1)}\nu^{(1)}\left[\frac{\partial}{\partial r}\frac{\Psi_2^{(1)}(r)}{r}\right]$$

$$-\rho^{(2)}\nu^{(2)}\left[\frac{\partial}{\partial r}\frac{\Psi_2^{(2)}(r)}{r}\right]=0. \tag{33}$$

The solutions of Eqs. (32) have the form

$$\Psi_1^1(r)=b_1r^m;\quad \Psi_1^2(r)=c_1r^{-(m+1)};$$

$$\Psi_2^1(r)=b_2\frac{i_m(x_1r)}{i_m(x_1)};\quad \Psi_2^2(r)=c_2\frac{k_m(x_2r)}{k_m(x_2)};$$

$$\Psi_3^1(r)=b_3\frac{i_m(x_1r)}{i_m(x_1)};\quad \Psi_3^2(r)=c_3\frac{k_m(x_2r)}{k_m(x_2)}, \tag{34}$$

where $x_1=\sqrt{S/\nu_1}$ and $x_2=\sqrt{S/\nu_2}$; $i_m(x)$ and $k_m(x)$ are modified Bessel functions of the first and third kinds; and, b_j and c_j are constants.

Solving the electric part of the problem, the equations (3) with the boundary conditions (4) and (11)–(13), we represent the perturbation of the electric-field pressure and the

electric components of the tangential components of the stress tensor (17), using the solutions (34), in the form

$$p_E \approx -\frac{Q^2}{4\pi g_2^2}\sum_{m,n}\frac{1}{L_m}\left\{g_4 Z - m(m+1)g_2 g_5 \times\left[b_1+\left(x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}+(m+1)\right)b_3\right]\right\}Y_{mn}\exp(St);$$

$$\Pi_{1\Theta}-\Pi_{2\Theta}\approx\frac{Q_2(\sigma_2-\sigma_1)}{4\pi g_2^2}\sum_{m,n}\frac{1}{L_m}\left\{\eta_2 Z - m(m+1)\times g_2(1-R_0^{(2l+1)})\left[b_1+\left(x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}+(m+1)\right)b_3\right]\right\}\frac{\partial Y_{mn}}{\partial\Theta}\exp(St);$$

$$\Pi_{1\varphi}-\Pi_{2\varphi}\approx\frac{Q_2(\sigma_2-\sigma_1)}{4\pi g_2^2}\sum_{m,n}\frac{1}{L_m}\left\{\eta_2 Z - m(m+1)\times g_2(1-R_0^{(2l+1)})\left[b_1+\left(x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}+(m+1)\right)b_3\right]\right\}\frac{1}{\sin\Theta}\frac{\partial Y_{mn}}{\partial\varphi}\exp(St);$$

$$Q\equiv\frac{\varphi_0 R_0 g_2}{[\sigma_1 R_0+\sigma_2(1-R_0)]};\quad \lambda_i\equiv S\varepsilon_i+4\pi\sigma_i;$$

$$\mu_i\equiv\lambda_i+m(m+1)D_s\varepsilon_i;$$

$$L_m(\sigma,\varepsilon,S)\equiv d(\mu_i)+m(m+1)Qb(1-R_0^{(2m+1)});$$

$$d(\mu_j)\equiv\mu_1 m+\mu_2(m+1)+(\mu_1-\mu_2)(m+1)R_0^{(2m+1)};$$

$$d(\varepsilon_j)\equiv\varepsilon_1 m+\varepsilon_2(m+1)+(\varepsilon_1-\varepsilon_2)(m+1)R_0^{(2m+1)};$$

$$d(\lambda_i)\equiv\lambda_1 m+\lambda_2(m+1)+(\lambda_2-\lambda_2)(m+1)R_0^{(2m+1)};$$

$$g_1\equiv(2\lambda_1-\lambda_2(m+1))\sigma_2+\lambda_2\sigma_1(m-1);$$

$$g_2\equiv\varepsilon_2\sigma_1-\varepsilon_1\sigma_2;$$

$$g_3\equiv\sigma_2 d(\lambda_i)+g_1[1-R_0^{(2m+1)}];$$

$$g_4\equiv\varepsilon_2\sigma_1\{(m+1)\eta_3-(m-1)\sigma_1[d(\mu_i)+m(m+1)\times Qb(1-R_0^{(2m+1)})]\}+\varepsilon_1\sigma_2\{[m+(m+1)R_0^{(2m+1)}]\eta_1-2\sigma_2 d(\mu_i)-\sigma_2 m(m+1)Qb[(m+2)+(m-1)R_0^{(2m+1)}]\};$$

$$g_5\equiv[(m+1)\varepsilon_2\sigma_1+m\varepsilon_1\sigma_2]-(m+1)g_2R_0^{(2m+1)};$$

$$h_1\equiv(2\varepsilon_1-(m+1)\varepsilon_2)+\varepsilon_2\sigma_1(m-1);$$

$$h_2\equiv h_1(1-R_0^{(2m+1)})+\sigma_2 d(\varepsilon_i);$$

$$\eta_1\equiv g_1+m(m+1)D_s h_1;\quad \eta_2\equiv g_3+m(m+1)D_s h_2; \tag{35}$$

Q is the total charge of the drop–medium interface in the equilibrium state.

We now substitute the solution for the radial functions $\Psi_j^{(i)}(r)$ (34) into the system of boundary conditions (33). Taking Eqs. (35) into account and using the recurrence relations for the modified spherical Bessel functions

$$\frac{i_m'(x)}{i_m(x)}=\frac{i_{m+1}(x)}{i_m(x)}+\frac{m}{x},$$

$$\frac{i_m''(x)}{i_m(x)}=-\frac{2}{x}\frac{i_{m+1}(x)}{i_m(x)}+\left[1+\frac{m(m-1)}{x^2}\right],$$

$$\frac{k_m'(x)}{k_m(x)}=-\frac{k_{m-1}(x)}{k_m(x)}-\frac{m+1}{x},$$

$$\frac{k_m''(x)}{k_m(x)}=\frac{2}{x}\frac{k_{m-1}(x)}{k_m(x)}+\left[1+\frac{(m+1)(m+2)}{x^2}\right],$$

after transformations we obtain a homogeneous system of linear equations for the coefficients b_1, c_1, b_3, c_3, and Z

$$b_1+b_3\left[x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}+(m+1)\right]-c_1+c_3\left[x_2\frac{k_{m-1}(x_2)}{k_m(x_2)}+m\right]=0;$$

$$\left[\rho^{(1)}\nu^{(1)}2(m-1)-\frac{Q^2}{4\pi}\frac{(\sigma_2-\sigma_1)}{g_2}\frac{m(m+1)}{L_m}\times(1-R_0^{(2m+1)})\right]b_1+\left[\rho^{(1)}\nu^{(1)}\left\{-2x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}+x_1^2+2(m-1)(m+1)\right\}-\frac{Q^2}{4\pi}\frac{(\sigma_2-\sigma_1)}{g_2}\frac{m(m+1)}{L_m}\times(1-R_0^{(2m+1)})\left\{x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}+(m+1)\right\}\right]b_3+\left\{\frac{Q^2}{4\pi}\frac{(\sigma_2-\sigma_1)}{g_2^2}\frac{1}{L_m}\eta_2\right\}Z+[\rho^{(2)}\nu^{(2)}2(m+2)]c_1-\rho^{(2)}\nu^{(2)}\times\left[2x_2\frac{k_{m-1}(x_2)}{k_m(x_2)}+x_2^2+2m(m+2)\right]c_3=0;$$

$$mb_1+m(m+1)b_3+(m+1)c_1-m(m+1)c_3=0;$$

$$mb_1+m(m+1)b_3-SZ=0;$$

$$\left(\rho^{(1)}(S+2\nu^{(1)}m(m-1))-\frac{Q^2}{4\pi}\frac{1}{g_2}\frac{m(m+1)}{L_m}g_5\right)b_1+\left(2\rho^{(1)}\nu^{(1)}m(m+1)\left[x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}+(m-1)\right]-\frac{Q^2}{4\pi}\frac{1}{g_2}\frac{m(m+1)}{L_m}g_5\left\{x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}+(m+1)\right\}\right)b_3-\rho^{(2)}(S+2\nu^{(2)}(m+1)(m+2))c_1+2\rho^{(2)}\times\nu^{(2)}m(m+1)\left[x_2\frac{k_{m-1}(x_2)}{k_m(x_2)}+(m+2)\right]c_3$$

$$+\left[(m-1)(m+2)+\frac{Q^2}{4\pi}\frac{1}{g_2^2}\frac{g_4}{L_m}\right]Z=0.$$

This system possesses a nontrivial solution only if the determinant constructed from the coefficients with unknowns b_1, c_1, b_3, c_3, and Z is zero:

$$\det a_{\mathbf{l}k}=0; \quad 1\leqslant \mathbf{l}, k\leqslant 5; \tag{36}$$

$$a_{11}=1; \quad a_{12}=-1; \quad a_{13}=f_m^{(1)}(x_1)+(m+1);$$

$$a_{14}=f_m^{(2)}(x_2)+m; \quad a_{15}=0;$$

$$a_{21}=2(m-1)\rho^{(1)}\nu^{(1)}-\frac{Q^2}{4\pi}\frac{(\sigma_2-\sigma_1)}{g_2}\frac{(m(m+1))^2}{L_m}\times(1-R_0^{(2m+1)});$$

$$a_{22}=2(m+2)\rho^{(2)}\nu^{(2)};$$

$$a_{23}=\rho^{(1)}\nu^{(1)}[-2f_m^{(1)}(x_1)+x_1^2+2(m-1)(m+1)]-\frac{Q^2}{4\pi}\frac{(\sigma_2-\sigma_1)}{g_2}\frac{m(m+1)}{L_m}(1-R_0^{(2m+1)})\times\{f_m^{(1)}(x_1)+(m+1)\};$$

$$a_{24}=-\rho^{(2)}\nu^{(2)}[2f_m^{(2)}(x_2)+x_2^2+2m(m+2)];$$

$$a_{25}=\frac{Q^2}{4\pi}\frac{(\sigma_2-\sigma_1)}{g_2^2}\frac{1}{L_m}\eta_2; \quad a_{31}=m;$$

$$a_{32}=(m+1); \quad a_{33}=m(m+1); \quad a_{34}=m(m+1);$$

$$a_{35}=0; \quad a_{41}=m; \quad a_{42}=0;$$

$$a_{43}=m(m+1); \quad a_{44}=0; \quad a_{45}=-S;$$

$$a_{51}=\rho^{(1)}[S+2m(m-1)\nu^{(1)}]-\frac{Q^2}{4\pi}\frac{1}{g_2}\frac{m(m+1)}{L_m}g_5;$$

$$a_{52}=-\rho^{(2)}S-2(m+1)(m+2)\rho^{(2)}\nu^{(2)};$$

$$a_{53}=2\rho^{(1)}\nu^{(1)}m(m+1)[f_m^{(1)}(x_1)+(m-1)]-\frac{Q^2}{4\pi}\frac{1}{g_2}\frac{m(m+1)}{L_m}g_5\{f_m^{(1)}(x_1)+(m+1)\};$$

$$a_{54}=\rho^{(2)}\nu^{(2)}2m(m+1)[f_m^{(2)}(x_2)+(m+2)];$$

$$a_{55}=(m-1)(m+2)+\frac{Q^2}{4\pi}\frac{1}{g_2^2}\frac{g^4}{L_m};$$

$$f_m^{(1)}(x_1)\equiv x_1\frac{i_{m+1}(x_1)}{i_m(x_1)}; \quad f_m^{(2)}(x_2)\equiv x_2\frac{k_{m-1}(x_2)}{k_m(x_2)}.$$

4. The dispersion relation (36) was analyzed numerically. The parameter $W\equiv\varepsilon_2\varphi_0^2/4\pi$ was used to characterize the intensity of the electric field (charge of the drop) at the interface between the media. For $\sigma_2=0$ this is the same as the definition of the Raleigh parameter[6,11,12] $W\equiv Q^2/4\pi\varepsilon_2$.

The radius of the in-drop spherical electrode was assumed to be the same $R_0=0.1$ in all calculations. According to Refs. 6 and 7, the presence of a solid spherical core for $R_0<0.5$ inside the drop has virtually no effect on the character of the hydrodynamic motions that arise; this ensures that the model used in the analytic derivation of Eqs. (36) is acceptable.

Numerical methods were used to investigate the behavior of the family of dependences of the real (ReS) and imaginary (ImS) components of the complex frequency on the conductivity σ_2 of the exterior medium for various values of the permittivity of the drop $10\leqslant\varepsilon_1\leqslant 80$. The curves presented in Figs. 1a and 1b were obtained for $n=2$ and $\varepsilon_2=2$. In Figs. 2a–2d the analogous curves were obtained for $\varepsilon_2=4$.

All calculations were performed with a supercritical (with respect to the Rayleigh value) parameter W: $W=1.1W_*$ ($W_*=n+2$; $n=2,3,4,\ldots$ — mode number).[4] The dimensionless kinematic viscosity of the drop and the exterior medium were assumed to be the same $\nu_1=\nu_2=\nu=0.02$; the exterior medium was assumed to be much less dense than the drop medium $\rho^{(1)}=1$ and $\rho^{(2)}=0.01$; in all calculations the dimensionless electric conductivity of a drop was assumed to be 1, $\sigma_1=1$; moreover, the charge-carrier mobility and the surface diffusion coefficient were assumed to be $b=0.01$ and $D_s=10^{-5}$, respectively.

According to the numerical calculations, the complete family of branches of the dispersion relation is infinite because the set of poloidal vortex motions which arise is infinite.[11] Calculations were performed to investigate the instability of the capillary motions of the liquid that are not associated with poloidal vortices.[11] However, for completeness, all branches of the dispersion relation which are realized in a prescribed range of the characteristic physical parameters are presented in the dependences.

Two types of aperiodic instability are observed in the dependences shown in Fig. 1a. A type-I aperiodic instability with $\varepsilon_1=40$ is represented by the branch *1* and is characterized by a growth rate decreasing monotonically with increasing σ_2 from the maximum value for $\sigma_2=0$, corresponding to the growth rate of the Rayleigh instability, to zero for $\sigma_2=0.005$. As ε_1 varies in the range $10\leqslant\varepsilon_1\leqslant 80$, the changes in the behavior of the branch *1* fall within the thickness of the line representing this curve in Fig. 1a.

A type-II instability is illustrated by the branches *2* ($\varepsilon_1=80$) and 3 ($\varepsilon_1=40$). It is observed in the range $\sigma_2>0.005$ and is characterized by the fact that its growth rate increases with σ_2, reaches a maximum, and decreases to zero as σ_2 increases further. There is a one-to-one correspondence between each value of ε_1 in the interval[40,80]* and a curve of the one-parameter family, formed by a continuous mapping of the branch *2* into the branch *3* as ε_1 decreases from 80 to 40. In what follows we shall call this character of the dependences, with ε_1 varying continuously in the interval bounded by its values for which the diagram of the branches is presented, intermediate (relative to the branches *2* and *3*). It is evident from Fig. 1a that the growth rate of the instability is determined in some range of values of σ_2 that does not contain $\sigma_2=0$; on this interval it reaches a maximum and decreases to zero. The magnitude of the maximum and the width of the interval within which instability is observed to decrease substantially as ε_1 decreases. The position of the maximum shifts rightward. For $10\leqslant\varepsilon_1<40$ the type-II

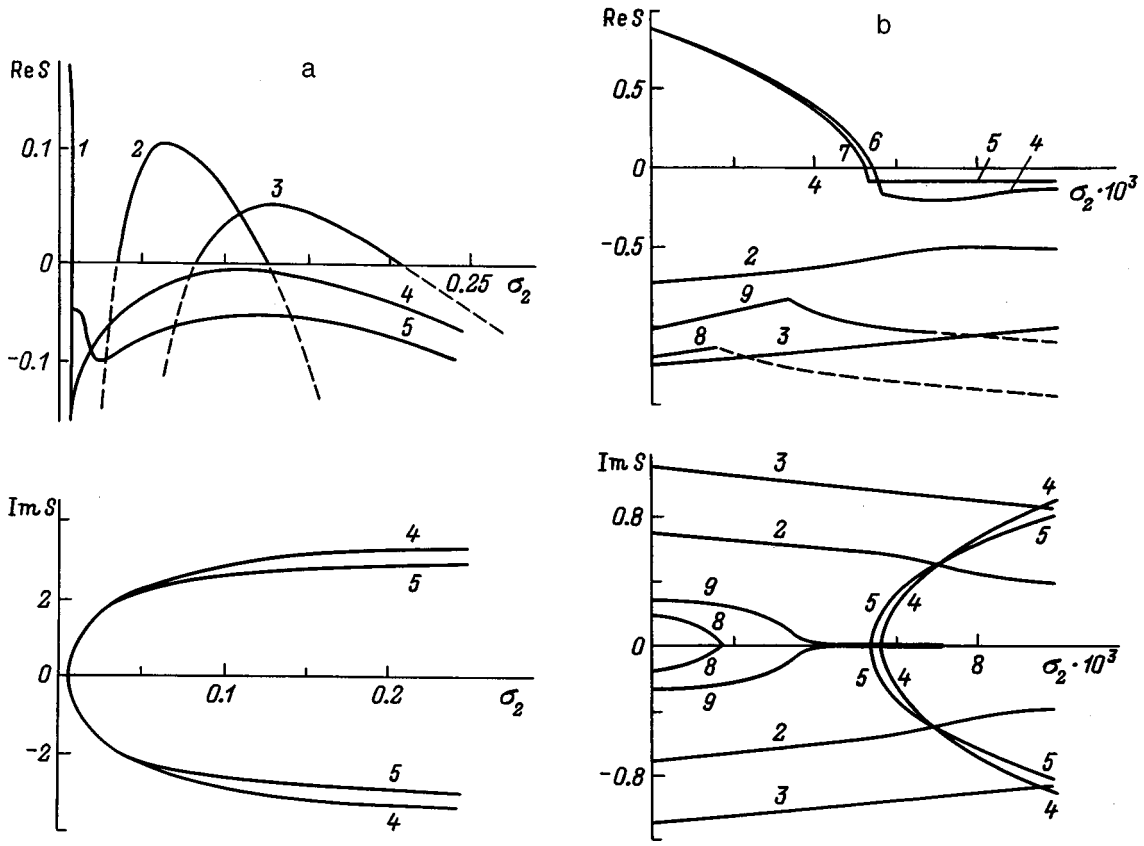

FIG. 1. a — Real and imaginary components of the dimensionless complex frequency of capillary motions of a liquid versus the dimensionless conductivity of the exterior medium: a branch corresponding to $\varepsilon_1=80$ lies within the thickness of the curve *1*, for $\varepsilon_1=80$ branch *2* of a type-II aperiodic instability and branch *4* of damped oscillatory motions are additionally observed; for $\varepsilon_1=40$ branch *3* of a type-II aperiodic instability and branch *5* of damped oscillatory motions are observed. b — The part of Fig. 1a near the origin of the coordinates: the branches *2*–*5* characterize the behavior of the same branches as in Fig. 1a; *6* and *7*, in contrast to Fig. 1a, divide into two individual curves; *8* and *9* — poloidal vortex motions with $\varepsilon_1=80$ and 40, respectively.

aperiodic instability is not observed, and the corresponding branch lies on the bottom sheet of the Riemann surface (RS), on which the dispersion relation (36) is defined, everywhere with the exception of the small region $\sigma_2\subset[0,0.02]$ where for this branch $\mathrm{Re}\,S<0$. The corresponding curves are not shown in Fig. 1a so as not to clutter the figure. Thus, the characteristics of a type-II instability depend strongly on the electric properties of the medium.

In all cases studied, the branch *1* is accompanied by a branch of the dispersion relation determined on the bottom sheet of the RS, which coincides with branch *1* at least to within the thickness of the line.As σ_2 increases, continuations of the branch *1* and the accompanying branch of the bottom sheet into the region $\mathrm{Re}\,S<0$ are bounded by the branch point, from which a branch of vibrational motion develops — the branches *4* and *5* in Fig. 1a. The permittivity $\varepsilon_1=80$ corresponds to the branch *4*, while $\varepsilon_1=40$ corresponds to the branch *5*. For ε_1 taking on values in the range $40<\varepsilon_1<80$, the behavior of the corresponding branches is qualitatively the same, and the branches themselves occupy positions between the curves *4* and *5*. The frequency of the vibrational motions grows very little with increasing ε_1 and increases rapidly with increasing σ_2. The damping decrement of the vibrational motion increases with ε_1 for values of σ_2 from a neighborhood of the branch point. As σ_2 increases, the damping decrement of the periodic motions decreases appreciably with increasing ε_1 and passes through a minimum at $\sigma_2\approx0.1$. In the range $10\leqslant\varepsilon_1<80$ the branch whose dependence on σ_2 is similar to the branches *4* and *5* is essentially identical to the branch *5* and is not shown in Fig. 1a.

The left side of Fig. 1a is constructed in Fig. 1b on a larger scale. The branch *1* separates near its right-hand end into two parts, designated by the numbers *6* ($\varepsilon_1=80$) and *7* ($\varepsilon_1=40$). The calculations showed that variation of ε_1 in the range $10<\varepsilon_1<80$ does not appreciably influence the growth rate of the type-I aperiodic instability. It was found that the motions corresponding to the branches *2* ($\varepsilon_1=80$) and *3* ($\varepsilon_1=40$) for small values of σ_2 are damped oscillations. In addition, as σ_2 increases, the damping decrements decrease to ≈0.02 and switch to the bottom sheet of the RS. As σ_2 increases further, each of the branches *2* and *3* approaches the branch point, which is the origin of the same branch of the aperiodic instability (Fig. 1a), accompanied by an aperiodic branch on the bottom sheet of the RS (this branch coincides with the curve on the first sheet to within the thickness of the line). As ε_1 decreases from 80 to 40, the frequency and damping decrement of the motions discussed above increase from values determined by the branch *2* to values corresponding to the branch *3*. Besides the branches enumerated in Fig. 1b, there are also branches of vibrational motions *8* (for $\varepsilon_1=80$) and *9* (for $\varepsilon_1=40$), describing

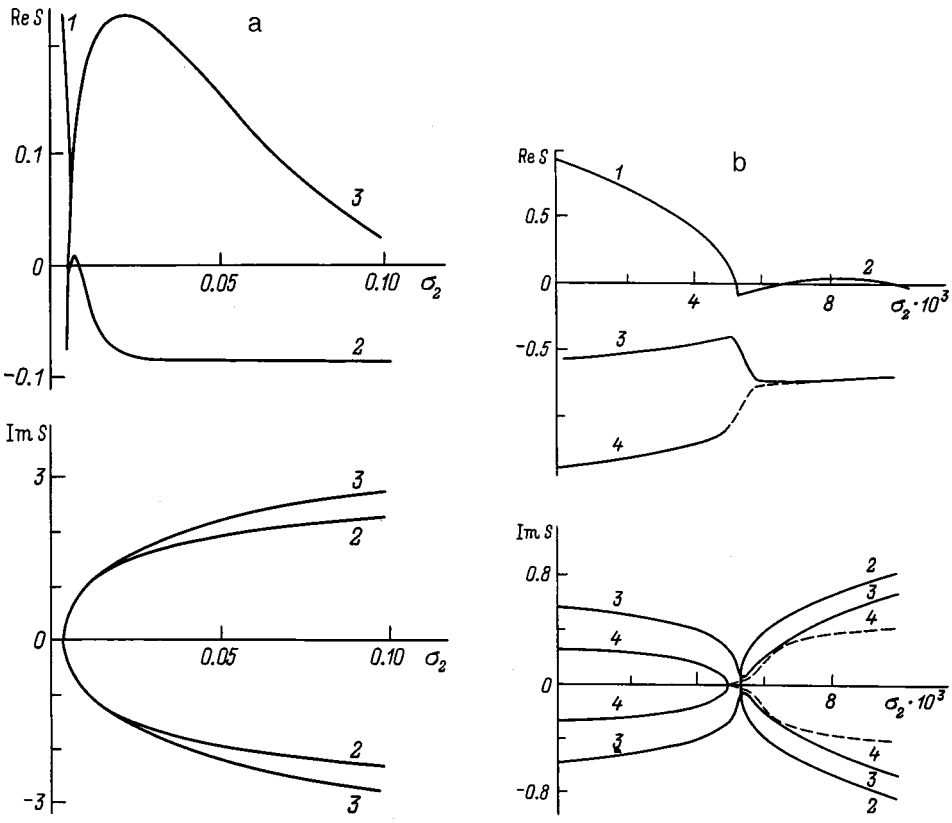

FIG. 2. a — Same as in Fig. 1a for $\varepsilon_2=4$. b — Behavior of the branches of Fig. 2a near the origin of coordinates calculated for $\varepsilon_1=80$: *2,3* — damped oscillatory motion. c — Behavior of the branches of the dispersion relation, calculated for $\varepsilon_1=70$ (*3,4*) and 60 (*5,6*), near the origin of the coordinates. d — Behavior of the branches of Fig. 2a with $\varepsilon_1=10$ near the origin of coordinates: *5* and *6* — damped oscillatory motions.

poloidal vortex motions. According to Fig. 1b, as ε_1 decreases from 80 to 40, the range of values σ_2 for which the corresponding motions are observable expands rightward. For $10\leqslant\varepsilon_1\leqslant40$ the configuration of the branches of the motions being analyzed is essentially identical to the figure corresponding to the value $\varepsilon_1=40$. The values $40\leqslant\varepsilon_1\leqslant80$ correspond to branches with intermediate behavior relative to the branches *2* and *3*, branches *4* and *5*, branches *6* and *7*, and branches *8* and *9*.

In contrast to Fig. 1a, the dependences presented in Fig. 2a were obtained for a different value of the permittivity of the exterior medium $\varepsilon_2=4$. The branch *1* has the same meaning and shows the same behavior as the same branch in Fig. 1a. However, there are fundamental differences. A type-II aperiodic instability is not observed, since the corresponding branches lie in the bottom sheet of the RS and are not shown in Fig. 2a. At the same time, the vibrational motion (denoted in Fig. 2a by the numbers *2* for $\varepsilon_1=80$ and *3* for $\varepsilon_1=10$) engendered at the branch point of branch *1* becomes unstable as σ_2 increases to $\sigma_2\approx0.01$. Thus, it is evident from Fig. 2a that for $\varepsilon_2=4$ and $10\leqslant\varepsilon_1\leqslant80$, in some range of values of σ_2 that does not contain $\sigma_2=0$ an oscillatory instability with growth rate reaching a maximum in this interval is observed. The value of the maximum and the width of the range of values of σ_2 where oscillatory instability occurs decrease substantially with increasing ε_1, which is opposite to the behavior of a type-II aperiodic instability in Fig. 1a. The position of the maximum shifts leftward as ε_1 increases. The frequency changes very little as a function of ε_1.

To describe the behavior of the branches of the dispersion relation for small values of σ_2 (on the left-hand side of Fig. 2a) in the range $10\leqslant\varepsilon_1\leqslant80$ it is not enough to consider only the extreme situations $\varepsilon_1=10$ and 80. For this reason, a configuration of branches calculated for $\varepsilon_1=80$ is presented in Fig. 2b; the configuration for $\varepsilon_1=70$ and 60 is presented in Fig. 2c; and, the configuration for $\varepsilon_1=10$ is presented in Fig. 2d. It is evident from Fig. 2b that for the scale employed, besides the unstable branches *1* and *2*, branches of two oscillatory motions *3* and *4* are present in the upper sheet. According to the numerical calculations, the branch *3* switches to the bottom sheet of the RS for $\sigma_2\gtrsim0.005$ and branch *4* for $\sigma_2\geqslant0.005$. In Fig. 2c, the previous enumeration with $\varepsilon_1=70$ is retained for the curves *3* and *4* (the unstable branches are once again described by the curves *1* and *2*). The same figure shows the computational results obtained with $\varepsilon_1=60$, and the branches associated with the instability correspond, to within the thickness of the line shown, to the branches *1* and *2*. The branches *5* and *6* form as the branches *3* and *4* reconnect as ε_1 decreases from 70 to 60. The branch *6* switches to the bottom sheet of the RS for $\sigma_2\gtrsim0.08$ and the branch *5* for $\sigma_2>0.008$. Further evolution of the branches *5* and *6* as ε_1 decreases to $\varepsilon_1=10$ reduces to a continuous increase in the growth rates of the instability and the frequencies of the motion 6 and to rightward expansion of the range of values of σ_2 where the motion 5 is determined. Figure 2d shows the corresponding dependences calculated for $\varepsilon_1=10$. For $\sigma_2\gtrsim0.005$ the frequency of the motions described by the branch *5* are much lower than the value of the corresponding damping decrement, making it possible to assume the motion to be aperiodically damped.

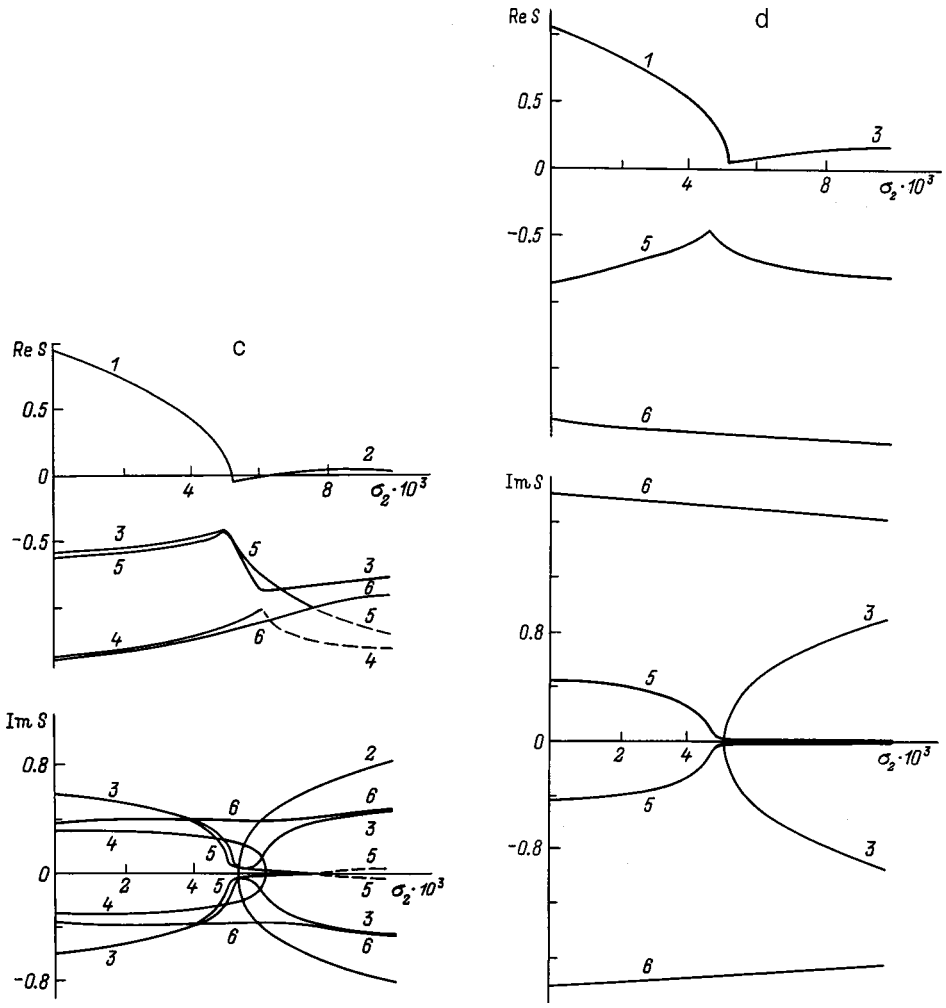

FIG. 2. (continued)

According to the numerical calculations, for $\varepsilon_1 = 10$ the branch *5* moves off the first sheet of the RS for $\sigma_2 \gtrsim 0.04$ and the branch *6* for $\sigma_2 \gtrsim 0.4$. The values $10 \leqslant \varepsilon_1 \leqslant 60$ correspond to branches with intermediate behavior relative to the same branches in Figs. 2c and 2d.

Figures 1a and 1b and 2a–2d show that for nonzero conductivity of the exterior medium, besides a Rayleigh instability, one other type of instability is also observed. Its character depends strongly on the dielectric properties of the exterior medium. Thus, as ε_2 varies from 2 to 4 the instability changes the aperiodic behavior to oscillatory. The aperiodic behavior is characterized by an increase in the growth rate of the instability with increasing ε_1, whereas the growth rate of the oscillatory instability decreases substantially with increasing ε_1. In both cases a non-Rayleigh instability is observed with a supercritical value of the parameter W and exists in a certain range of values of σ_2 not containing $\sigma_2 = 0$. For $\sigma_2 > 0.25$ instability was not observed in any of the situations that were analyzed. The calculations show that as ε_2 increases, the spectrum of the damped poloidal oscillations occurring together with the unstable motions becomes more complicated. In contrast to the case $\varepsilon_2 = 2$, for $\varepsilon_2 = 4$ variation of ε_1 leads to reconnection of certain branches of the poloidal oscillations.

Dependences similar to those considered above can be obtained for other modes. Figure 3 shows the unstable part of the branches associated with the type-II aperiodic instability, when $n = 4$. The branch *2* was constructed for $\varepsilon_1 = 80$, while

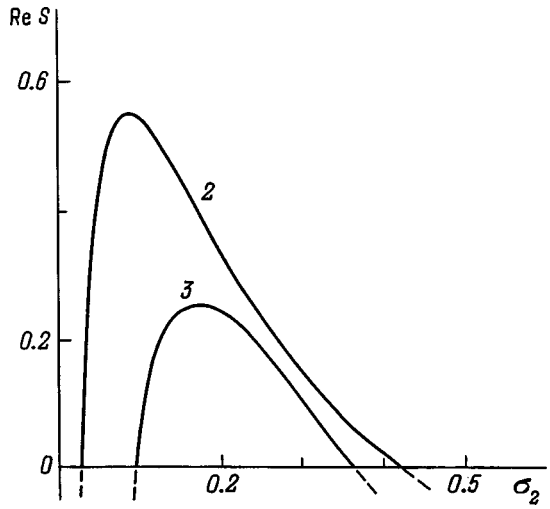

FIG. 3. The parts of the branches associated with a type-II aperiodic instability for $n = 4$: $\varepsilon_1 = 80$ (*2*) and $\varepsilon_2 = 30$ (*3*).

the branch *3* was constructed for $\varepsilon_1 = 30$. The instability under consideration is not observed even for $\varepsilon_1 = 20$. The calculations show that as the mode number increases, the growth rate of the type-II instability increases and the range of σ_2 and ε_1 for which this instability exists expands substantially. Comparing Figs. 1a and 3 illustrates this fact. It was found that the growth rate of the Rayleigh instability for $n = 4$ rapidly decreases to zero as σ_2 increases to ≈ 0.005, it is virtually independent of ε_1, and on the scale of Fig. 3 it merges with the ReS axis. Qualitatively, however, the branches of the dispersion relation with $n = 4$ behave similarly to the branches considered in Figs. 1a and 1b.

CONCLUSIONS

A charged drop of electrically conducting liquid in an electrically conducting medium is subject to several types of instabilities: two aperiodic, Rayleigh-type[12] (whose growth rates depend differently on the electric conductivity of the exterior medium), and oscillatory. The oscillatory instability arises when the position of the maximum electric-field pressure runs ahead in phase (as a result of the finiteness of the rate of charge redistribution during capillary oscillations) compared with the maxima of the corresponding capillary waves. Only one type of instability can occur for any specific value of σ_2.

*Refs. 40 and 80 are cited in the Russian original, but are not listed in the reference section.

[1] A. G. Baily, Sci. Prog. Oxf. **61**, 555 (1974).
[2] V. I. Kozhenkov and N. A. Fuks, Usp. Khim. **45**, 2274 (1976).
[3] A. I. Grigor'ev, Yu. V. Syshchikov, and S. O. Shiryaeva, Zh. Prikl. Spektrosk. **62**, 2020 (1989).
[4] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Ser. Fiz. , No. 3, 3 (1994).
[5] A. I. Grigor'ev, D. F. Belonozhko, S. O. Shiryaeva, and M. I. Munichev, Pis'ma Zh. Tekh. Fiz. **23**(16), 38 (1997) [Tech. Phys. Lett. **23**, 631 (1997)].
[6] A. I. Grigor'ev, S. O. Shiryaeva, V. A. Koromyslov, and M. I. Munichev, Pis'ma Zh. Tekh. Fiz. **22**(10), 23 (1996) [Tech. Phys. Lett. **22**, 398 (1996)].
[7] A. I. Grigor'ev, V. A. Koromyslov, and S. O. Shiryaeva, Zh. Tekh. Fiz. **67**(9), 8 (1997) [Tech. Phys. **42**, 991 (1997)].
[8] J. R. Melcher and C. V. Śmith, Phys. Z. **12**, 778 (1969).
[9] A. I. Grigor'ev and A. É. Lazaryants, Zh. Tekh. Fiz. **63**(10), 12 (1993) [Tech. Phys. **38**, 840 (1993)].
[10] S. O. Shiryaeva, A. I. Grigor'ev, A. É. Lazaryants *et al.*, IMRAN Preprint No. 17 [in Russian], Institute of Microelectronics, Russian Academy of Sciences, Yaroslavl (1994).
[11] S. O. Shiryaeva, M. I. Minuchev, and A. I. Grigor'ev, Zh. Tekh. Fiz. **66**(7), 1 (1996) [Tech. Phys. **41**, 635 (1996)].
[12] Baron Rayleigh (J. W. Strutt), Philos. Mag. **14**, 184 (1982).

Translated by M. E. Alferieff

Physical model for the evolution of the defect system of silicon carbide with allowance for the internal elastic stress fields during implantation of Al^+ and N^+ and subsequent annealing

D. V. Kulikov, Yu. V. Trushin, P. V. Rybin, and V. S. Kharlamov

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted November 20, 1998; resubmitted April 30, 1999)
Zh. Tekh. Fiz. **69**, 43–50 (October 1999)

A theoretical analysis is offered for the formation and development of defects in silicon carbide implanted with nitrogen and aluminum ions and then annealed. The diffusion of defects, the formation of complexes, and the influence of the internal elastic stress fields produced by the implanted ions and the created complexes on the migration of interstials are taken into account. The computed distributions of defects agree satisfactorily with the experimental data. Certain kinetic parameters of silicon carbide are estimated numerically. © *1999 American Institute of Physics.* [S1063-7842(99)00610-8]

INTRODUCTION

Silicon carbide (SiC) and materials based on it are promising for high-power, high-frequency electronics. SiC is a wide-gap semiconductor, which could find application in electronic devices operating at high temperatures and in the presence of hard irradiation and a corrosive environment.[1–5] The promise of silicon carbide for applications was discovered back in the 1950s and 1960s.[6] However, scientific and industrial interest in SiC waned because at that time it was impossible to produce high-quality SiC crystals and also because of the rapid development of silicon electronics. Recent progress in growing bulk single crystals and thin films of SiC has given a new impetus for extensive investigation of this material.[7,8]

A characteristic feature of silicon carbide is the large number of structural forms, or polytypes, of this material,[9,10] which, however, all have approximately the same density (3.2 g/cm^2 or 9.66×10^{23} at./cm^3).[11] The arrangement of the nearest-neighbor atoms is identical in all forms, only the packing sequence of the carbon–silicon double layers changes.[12,13] There are more than 250 different polytypes,[14] but only two are investigated in connection with applications in electronic devices: the cubic form (3C) and the hexagonal form (6H).[12] The 3C form is interesting because it has the highest electron mobility (1000 cm^2/V·s) of all the polytypes, and the 6H structure has the largest band gap — 2.9 eV.[12]

The diffusion of intrinsic defects in SiC has been studied by many authors, both experimentally[15] and theoretically.[16–18] However, there are few quantitative data on the diffusion parameters of the material, and the results that have been obtained differ strongly. For example, the migration activation energies of the intrinsic atoms vary from 1.47[18] and 3.26[15] to 7–8 eV[16] for carbon and from 6[18] to 9 eV[16] for silicon. For vacancies, the only result available was obtained by molecular-dynamics calculations:[18] the migration activation energies are 6.1 eV for a carbon vacancy and 7.4 eV for a silicon vacancy.

Controllable and selective doping is required in order to use SiC in semiconductor devices. The most widely used dopants are N and P for *n*-type and B, Al, and Ga for *p*-type.[13] However, diffusion doping of SiC is difficult because of the smallness of the impurity diffusion coefficients, which for most impurities range from 10^{-13} to 10^{-8} cm^2/s for the temperature range 1700–2400 °C.[19] The activation energies for the migration of Al and N atoms are 6.1[19] and 7.6 eV[20], respectively. Therefore, satisfactory diffusion rates require temperatures above 2000 °C, at which not only does the oxide mask ordinarily used evaporate, but the SiC itself can decompose chemically. An alternative path is ion implantation. Its advantage is that the impurity distributions and concentrations can be controlled by varying the energy and dose of the implanted ions. Here problems associated with radiation damage to the material arise. However, they can be solved by annealing the irradiated sample at lower temperatures and for shorter times than for diffusion doping. The annealing regime required for the optimum result depends on the types and concentrations of defects generated by irradiation. For this reason, it is necessary to study the processes occurring during irradiation and annealing of SiC. Both experimental[21–35] and theoretical.[36–40] investigations of this kind have been performed for different types of ions.

The theoretical investigations have consisted in simulations of the ranges of the ions in irradiated SiC using the TRIM code or other computer codes based on the Monte Carlo method.[40] As a result, it has been found that dynamical and channeling effects during ion irradiation must be taken into account to describe the experimental data correctly.[36–39] Moreover, it has been found that the particular polytype that is irradiated plays a large role in implantation processes.[40]

The main experimental results are as follows. Depending on the mass of the atom and the irradiation temperature, the defect density increases with increasing ion dose, and at a certain critical value of the energy per unit volume transferred to nuclear processes, an amorphized layer forms. At room temperature this critical energy density is approxi-

mately 2×10^{21} keV/cm^3 (Refs. 21 and 23), which corresponds to a defect level of 0.2–0.25 displacements per atom.[24] During implantation the properties of the material change. At low ion doses the strength increases, while at high doses it decreases with the onset of amorphization.[24,26] The chemical activity and the optical and electrical properties of SiC also change.[27,28] For use of the material in electronic devices the defect density must be lowered. For weakly damaged unamorphized material virtually all of the defects generated are annealed at temperatures 200–1000 °C.[22,25] The behavior of amorphized layers during annealing is much more complicated and has not yet been adequately studied. For example, it is reported in Refs. 22 and 25 that recrystallization of amorphized SiC starts at temperatures above 1450 °C, whereas in Ref. 29 the onset of recrystallization was observed at 1000 °C. In Refs. 31–33, annealing of SiC samples irradiated with 200-keV Ge^+ ions was investigated at temperatures of 500, 950, and 1500 °C by various methods. It was found that at a dose of 1×10^{14} cm^{-2} a large number of defects do not vanish during annealing at 1500 °C, although the material does recrystallize, while for 1×10^{15} cm^{-2} the material contains regions of high defect density after annealing at 1500 °C. However, some recrystallization of the material is observed. A polytypic transition 6H → 3C at temperatures >1300 °C was reported. Since complete annealing of amorphous SiC has still not been achieved, many authors strive to prevent amorphization by increasing the irradiation temperature to 1800 °C (see, for example, Refs. 11, 34, and 35). For example, the authors of Ref. 35 showed that implantation of Ga and Sb ions at temperatures 200–400 °C suppresses amorphization. However, it has not been reported in any work that amorphization was completely prevented. It has been shown that the minimum density of defects is observed at temperatures 600–1000 °C.[34,35] This is explained by the fact that at low temperatures the recombination rate of vacancies and interstitials is too low to prevent their accumulation, while at high temperatures radiation-stimulated segregation and cluster formation lead to the appearance of secondary defects. Moreover, surface degradation is possible during annealing at temperatures above 1000 °C.[35]

At present, investigators are directing their efforts toward developing a technology that would make it possible to produce SiC-based devices by an industrial method.[8,13,41] These investigations have made it clear that the material properties required for using SiC impede its production. For example, the high-temperature stability is related with the low mobility of the atoms. Therefore, as we have said, extremely high temperatures are necessary to stimulate diffusion and annealing processes. However, processes occurring at high temperatures are difficult to monitor and greatly complicate the device fabrication technology. Moreover, material contamination and degradation problems arise in connection with high-temperature processes. Therefore it is very important to develop a low-temperature technology.

Ion implantation is a standard process for semiconductor technology. For example, ion implantation can be used to produce $(SiC)_{1-x}(AlN)_x$ — a quasibinary system widely

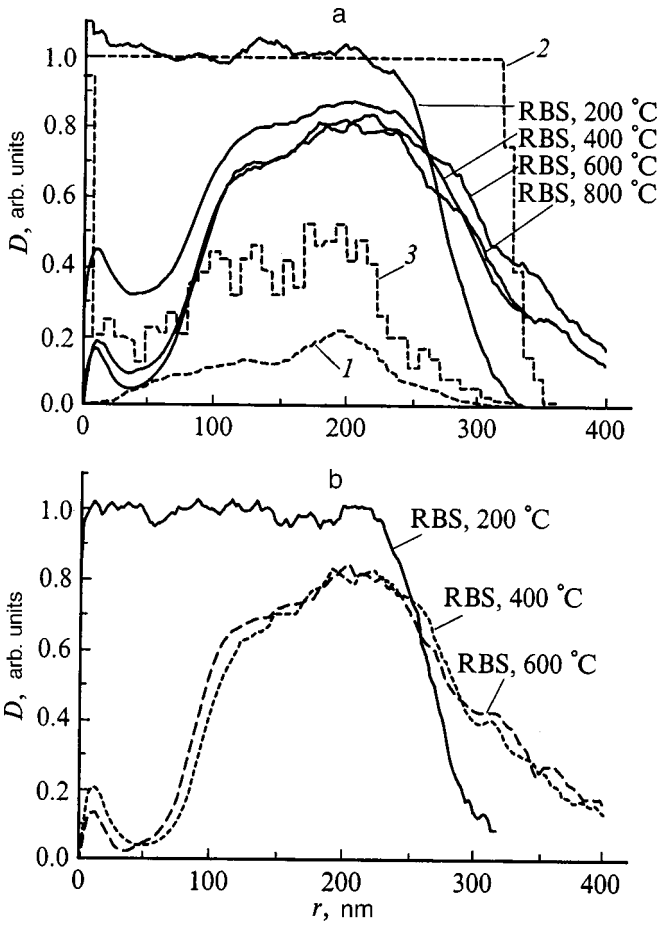

FIG. 1. a — Comparison of experimental and computed ballistic characteristics of implanted silicon carbide (D — defect density; r — depth): experimental RBS depth profiles of defects for various substrate temperatures; *1* — computed profile of implanted nitrogen and aluminum ions, *2* — computed total profile of defects taking into account only ballistic (cascade) processes, *3* — computed ballistic profile of defects taking into account the total athermal recombination. b — Experimental RBS depth profiles of defects after annealing in samples irradiated at various substrate temperatures and then annealed for 1 h at 1200 °C.

studied in the last few years. This is motivated by the qualities of the system, such as, first and foremost, the good miscibility of silicon carbide and aluminum nitride and, second, the good correspondence of their crystal lattices as well as the possibility of varying the band gap from 2.9 (6H–SiC) to 6.2 eV (2H–AlN).[42–45] The use of combined implantation of Al^+ and N^+ ions in 6H–SiC at high temperatures to obtain $(SiC)_{1-x}(AlN)_x$ was first reported in Ref. 46. Each sample was irradiated with ions in the following order: first with 65-keV nitrogen ions with dose 5×10^{16} cm^{-2}, then with 120-keV nitrogen ions with dose 1.3×10^{17} cm^{-2}, and finally with 100-keV aluminum ions with dose 5×10^{16} cm^{-2} and then with energy 160 keV and dose 1.3×10^{17} cm^{-2}. The substrate temperatures were 200, 400, 600, and 800 °C. After implantation the samples were investigated by the Rutherford backscattering method (RBS/C) using a 1.4 meV He^+ ion beam. The RBS spectra were analyzed using a specially developed computer program, which made it possible to obtain the depth distributions of the structural defects[47] displayed in Fig. 1a. Then the samples were annealed at

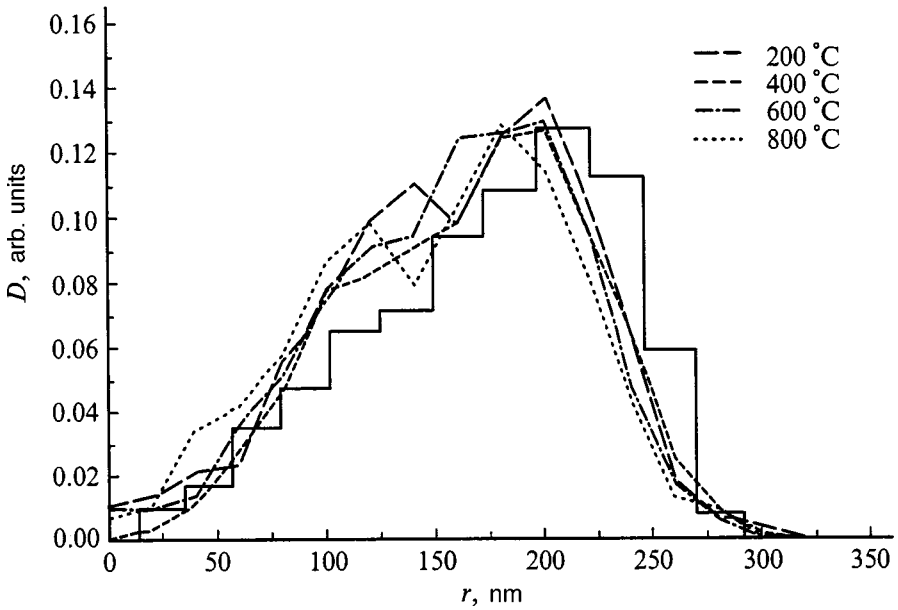

FIG. 2. Nitrogen depth profiles for different substrate temperatures: solid line — computed (using DYTRIRS program) ballistic nitrogen depth profile taking into account all subsequent irradiations; broken lines — experimental ERD nitrogen distributions for various irradiation temperatures.

1200 °C for 1 h and once again investigated by RBS (Fig. 1b). It was found that only irradiation at 200 °C gives a completely amorphized layer, while irradiation at other temperatures increases the density of defects but complete amorphization is not observed. The polytypic transformations 6H–SiC→ 3C–SiC was observed at depths 50–120 nm in the sample irradiated at 800 °C and was identified using the RBS data and the results of transmission electron microscopy.[48–50] In Ref. 47 it is concluded that the solid solution SiC/AlN forms in the irradiated sample. In the experimental part of Ref. 48, the nitrogen ion distributions presented in Fig. 2 (the experimental conditions are the same as in Refs. 46 and 47) were determined by the ERD method for silicon carbide implanted with Al^+ and N^+ ions. In Ref. 51 elastic compression stress fields at depths greater than 100 nm in nitrogen and aluminum ion implanted SiC samples (the experimental conditions are the same as in Refs. 46–50) were observed by the PIRR method (polarized infrared spectroscopy).

PHYSICAL MODEL

The ballistic distributions of N^+ and Al^+ ions (DYTRIRS) obtained in Refs. 48–50 are presented in Fig. 1a (curve *1*). Summation (taking the dose into account and neglecting recombination) of the depth distributions of defects calculated with the TRIRS program for a single incident ion[48–50] gives curve *2* in Fig. 1a. The ballistic distributions of implanted nitrogen (as a result of all four successive irradiations with the doses used in the experiment) are presented in Fig. 2 (DYTRIRS). Comparing the ballistic data (curve *2* in Fig. 1a) and the RBS results for irradiation at 200 °C shows that this temperature does not lead to annealing of the modified (''amorphous'') material at the depth of the implanted nitrogen and aluminum ions (curve *1*).

If it is assumed, for a rough estimate, that all vacancies and interstitial atoms separated from one another at spontaneous recombination distances (see, for example, Ref. 52) recombine athermally, then the remaining intrinsic defects and implanted ions will give a total profile lying below the experimental RBS data (curve *3* in Fig. 1a). Substantial differences are already seen from the two limiting cases (only ballistic distributions — curve *2*, and with the total recombination taken into account — curve *3* in Fig. 1a), and hence their comparison with the experimental data (RBS)[47–50] reveals the role of diffusion processes accompanying high-temperature implantation in regions of irradiated SiC at different depths.

In the experiments (ERD) whose results were first presented in Ref. 48, the N^+ depth profiles were measured (Fig. 2, ERD) in SiC samples. Comparing them with the ballistic computed nitrogen profile for all irradiation sequences at the corresponding experimental doses (Fig. 2, DYTRIRS) shows that the implanted nitrogen essentially does not diffuse at the temperatures employed.

In accordance with Fig. 1a (Refs. 46 and 47) and also Refs. 53 and 54, to simplify the description of the physical changes occurring, we shall separate in the sample two large regions according to depth: $r<120$ nm and $r>120$ nm. This separation was introduced because 1) in the region $r>120$ nm the distributions of defects after irradiation at temperatures 400, 600, and 800 °C behave approximately the same way: for $120<r<270$ nm the dependence on r can be neglected; for $r>270$ nm the density of RBS scattering centers decreases virtually identically for samples irradiated at different temperatures; 2) in the region $r<120$ nm all three irradiated samples behave differently, which is reflected in the defect distributions.

In Ref. 53 we proposed a model that describes the evolution of defects in SiC for the experimental conditions of Refs. 46–50 only in the region $r>120$ nm. The surface region was not studied in Ref. 53.

In elaborating the model of Ref. 53, to describe the processes occurring when SiC is irradiated with Al^+ and N^+ ions at temperatures 400 and 600 °C in the entire depth range, we propose the following (Ref. 54).[1)]

1. The Al^+ and N^+ ions, interacting with the crystal lattice of SiC, form interstials (i) and vacancies (v) in both subsystems of the material (carbon — C and silicon — S) (the relative concentrations are $C_{iC}(r,t)$, $C_{vC}(r,t)$, $C_{iS}(r,t)$, and $C_{vS}(r,t)$ respectively).

2. Implanted Al^+ ions occupy vacancies in the silicon sublattice, and N^+ ions occupy vacancies in the carbon sublattice[5] [the relative total density of both ions is $C(r,t)$].

3. Diffusion of interstitial carbon (iC) and silicon (iS) atoms and their recombination with their own vacancies (vC and vS) are possible during irradiation.

4. Vacancies and implanted ions are assumed to be stationary, since a) vacancies have a higher migration activation energy[18] and b) comparing the experimental[47–49] and computed (using DYTRIRS[48,49]) depth profiles of the N^+ ions in SiC (Fig. 2) shows that these ions do not diffuse at the temperatures employed (this is also assumed for Al^+ ions).

5. In the region $r>120$ nm the carbon (iC) and silicon (iS) interstitials can form complexes consisting of two carbon or silicon interstitials [the relative concentrations are $C_{2C}(r,t)$ and $C_{2S}(r,t)$].

6. Complexes of interstitials do not form in the region $r<120$ nm. This can be explained by the effect of the surface, specifically, impurities on the surface (see peak in the distributions of defects for $r<50$ nm, Fig. 1a).

7. At depths of the order of 100 nm, i.e., where the defect density increases (Fig. 1), there exist internal elastic compression stress fields which are produced by the implanted ions and by the complexes which are formed. Such internal stress fields, identified by PIRR,[51] influence the diffusion of interstials. In Ref. 53, this influence was taken into account effectively by decreasing the diffusion coefficient in the region near 100 nm.

On the basis of the assumptions (1)–(7) we write the following system of kinetic balance equations for the abovenamed relative concentrations of point defects, which depend on depth r and time t:

$$\frac{\partial C_{iC}(r,t)}{\partial t}=g_{iC}(r,t)-\operatorname{div}J_{iC}(r,t)-\mu_C D_{iC}C_{iC}(r,t)C_{vC}(r,t)-\alpha_{iC}D_{iC}C_{iC}^2(r,t), \tag{1}$$

$$\frac{\partial C_{vC}(r,t)}{\partial t}=g_{vC}(r,t)-g_N(r,t)-\mu_C D_{iC}C_{iC}(r,t)C_{vC}(r,t), \tag{2}$$

$$\frac{\partial C_{iS}(r,t)}{\partial t}=g_{iS}(r,t)-\operatorname{div}J_{iS}(r,t)-\mu_S D_{iS}C_{iS}(r,t)C_{vS}(r,t)-\alpha_{iS}D_{iS}C_{iS}^2(r,t), \tag{3}$$

$$\frac{\partial C_{vS}(r,t)}{\partial t}=g_{vS}(r,t)-g_{Al}(r,t)-\mu_S D_{iS}C_{iS}(r,t)C_{vS}(r,t), \tag{4}$$

$$\frac{\partial C_{2C}(r,t)}{\partial t}=\alpha_{iC}D_{iC}C_{iC}^2(r,t), \tag{5}$$

$$\frac{\partial C_{2S}(r,t)}{\partial t}=\alpha_{iS}D_{iS}C_{iS}^2(r,t), \tag{6}$$

$$\frac{\partial C(r,t)}{\partial t}=g_N(r,t)+g_{Al}(r,t). \tag{7}$$

Here $g_l(r,t)$ is the rate of generation of the corresponding defects ($l=i$C, iS, vC, vS, Al, N), and under irradiation $g_l(r,t)=JP_l(r)$, where the flux density of the incident ions is $J=6.24\ \mathrm{cm^{-2}\cdot s^{-1}}$,[46,47] and $P_l(r)$ is the depth distribution of defects l produced by one incident ion, which are determined from TRIRS calculations[48,49] (the total profiles for all defects taking the dose into account and neglecting recombination are presented in Ref. 48 (Fig. 1a, curve *2*); μ_λ is the recombination parameter for vacancies and interstials in the subsystem $\lambda=$ C, S; $D_j=D_{0j}\exp(-\varepsilon_j^m/kT)$ and ε_j^m are the diffusion coefficient and the migration activation energy of interstitial carbon or silicon atoms ($j=i$C, iS), $\varepsilon_{iC}^m=1.47$ eV,[18] and since the value of ε_{iS}^m is unknown from the literature, it was varied during the calculations; $D_{0j}=10^{-3}\ \mathrm{cm^2\cdot s^{-1}}$; α_j is a parameter describing the formation of complexes of the corresponding defects ($\alpha_j\approx 4\pi a$ for $r>120$ nm and $\alpha_j=0$ for $r<120$ nm (according to the assumptions 5 and 6) ($j=i$C, iS), where $a\approx 2$ Å is the average interatomic distance in SiC).

The effect of elastic stress fields on the diffusion of interstitials is taken into account by introducing a drift term in the expression for the flux of migrating defects $J_j(r,t)$ ($j=i$C, iS) (see Eqs. (1) and (3))[52]

$$J_j(r,t)=-D_j\nabla C_i(r,t)-D_j\frac{C_j(r,t)}{kT}\nabla E_{str}^j(r,t), \tag{8}$$

where $E_{str}^j(r)$ is the interaction energy of a defect j interacting with the stress field.

Only qualitative data on the spatial dependence $E_{str}^j(r)$ are available (see Table 1 in Ref. 51), but this dependence is not known accurately, so that various methods were used to calculate the expression for the flux. In Ref. 53 the fact that the diffusion of carbon and silicon interstitials in the region $r<120$ nm is impeded by possible compression stresses (see assumption 7) was taken into account effectively in the diffusion terms $D_j\Delta C_j(r,t)$ in Eqs. (1) and (3) by decreasing the diffusion coefficients without considering the drift term in the expression for the divergence of the flux. In addition, such a decrease was prescribed at different points of the region $r<120$ nm differently, since the stress field should depend on the spatial coordinate, as is evident from the inhomogeneity of the defect distribution in Fig. 1a. Moreover, in Ref. 53, the decrease of the diffusion coefficient of interstials was prescribed as follows: the diffusion coefficient of the interstitials in the direction of decreasing field (i.e., from the volume into the surface region) is greater than in the reverse direction.

To refine the results obtained, in the present work an attempt was made to take into account the drift term in the

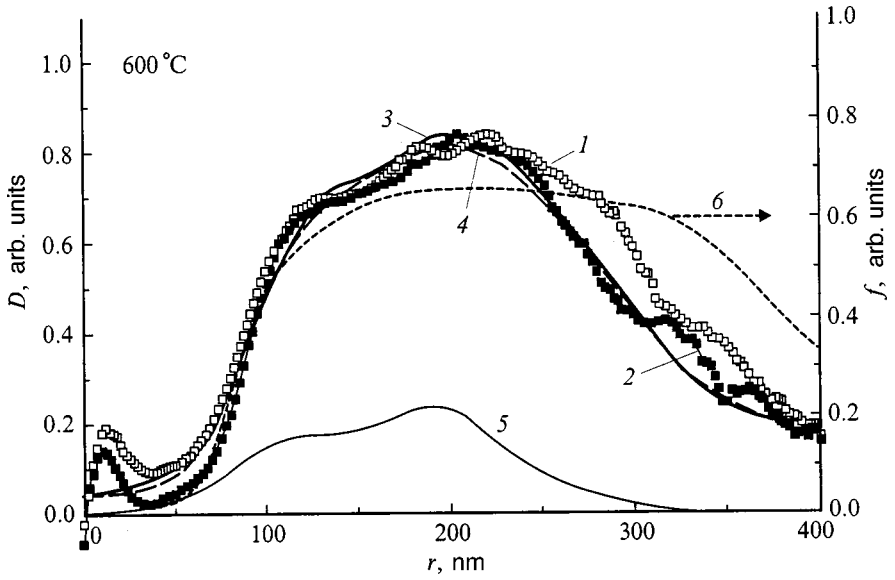

FIG. 3. Comparison of the experimental RBS depth profiles of defects[47–49] with the computed total defect distributions (taking diffusion processes into account) in SiC samples irradiated at 600 °C and annealed at 1200 °C: *1,2* — experimental results of Refs. 47–49 on irradiation and annealing, respectively; *3* — calculation (the irradiation and annealing profiles are identical), *4* — computational results from Ref. 53; *5* — the total ballistic depth profile of ions in SiC, calculated using DYTRIRS;[48] *6* — spatial dependence of the interaction energy $E^j_{\text{str}}(r)$ of defect j interacting with the stress field, i.e., the function $f(r)$.

expression for the flux of migrating interstitials (8). For this, the interaction energy $E^j_{\text{str}}(r,T)$ of interstitials interacting with the stress field was represented as

$$E^j_{\text{str}}(r,T)=\Delta\Omega_j \,\text{Tr}\,\sigma(r), \tag{9}$$

where $\Delta\Omega$ is the defect-induced change in the atomic volume and $\text{Tr}\,\sigma(r)$ is the trace of the stress field $\sigma_{ij}(r)$.

To calculate $\sigma_{ij}(r)$ it is necessary to sum the stress fields from the implanted ions and the complexes formed, which make their own contribution to the resulting field. Since this is extremely difficult to do for the present problem, the following approximation was proposed:

$$\text{Tr}\,\sigma(r)=\frac{A}{a^3}f(r), \tag{10}$$

where a is the interatomic parameter, A is a factor with the dimensions of energy, and $f(r)$ is a dimensionless function of the coordinates, which gives the spatial dependence $E^j_{\text{str}}(r,T)$.

Since the internal stresses should reflect the defect pattern that determines them, the function $f(r)$ was chosen in the form of the relative distribution of defects in a sample implanted with ions at 600 °C and annealed, i.e., actually as the distribution of stable defects (curve *6* in Fig. 3). Then, substituting expression (10) into Eq. (9), we obtain

$$E^j_{\text{str}}(r,T)=K(T)f(r), \tag{11}$$

where the factor $K(T)=\Delta\Omega_j A/a^3$ determines the interaction energy.

The temperature dependence of the energy $E^j_{\text{str}}(r,T)$ can be explained by the fact that defects of different nature are formed at different irradiation temperatures and create their own stress fields. This approximation makes it possible to solve correctly Eqs. (1) and (3) taking into account the drift term in the expression for the flux $J_j(r,t)$ ($j=i$C, iS; see Eq. (8)).

Since it is assumed that there are no defects in the initial (before irradiation) samples, the initial conditions for the concentrations will be

$$C_l(r,t)\big|_{t=0}=0, \tag{12}$$

where $l=i$C, iS, vC, vS, Al, N.

The boundary conditions for mobile defects $j=i$C, iS were chosen as

$$\left.\frac{\partial C_j(r,t)}{\partial r}\right|_{r=0}=0, \tag{13}$$

i.e., no flux of diffusing defects through the surface, and

$$C_j(r,t)\big|_{r=\infty}=0, \tag{12a}$$

i.e., the concentrations of defects $j=i$C, iS is zero at infinite depth.

To describe annealing of the irradiated samples at 1200 °C, we assumed that 8) recombination of vacancies and interstitials is possible, 9) new complexes do not form, and 10) the complexes formed during irradiation are assumed to be stable.

In accordance with the assumption 1, during annealing $g_l(r,t)=0$ for all types of defects ($l=i$C, iS, vC, vS, Al, N), $\alpha_j=0$ ($j=i$C, iS; see assumption 9), i.e., only Eqs. (1)–(4) remain in the system (1)–(7).

RESULTS

The system of equations (1)–(7), describing the evolution of defects in SiC irradiated at 400 and 600 °C, and the system (1)–(4) for describing annealing of a sample at 1200 °C were solved numerically using the MGEAR program.[55] The result was the distribution of the concentrations of defects in SiC after implantation at 400 and 600 °C and annealing of irradiated samples at 1200 °C. Figure 3 shows the computed and experimental (RBS) total distributions of defects after implantation. It is evident that the the-

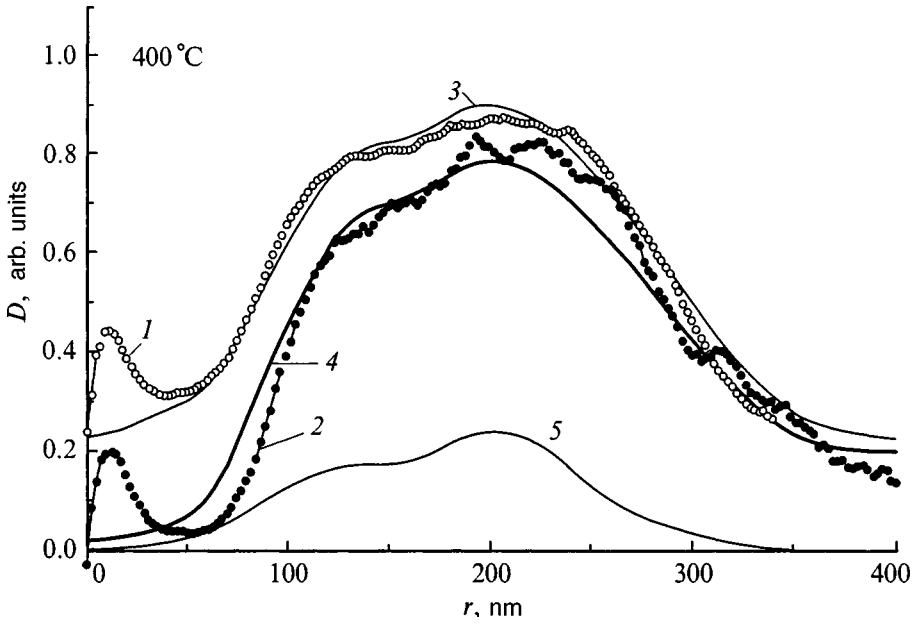

FIG. 4. Same as in Fig. 3 but the samples were irradiated at 400 °C (the curves *1–5* correspond to the curves in Fig. 3).

oretical results (curves *3* and *4*) agree satisfactorily with the experimental data for irradiation (curve *1*) and for annealing (curve *2*).

Figure 4 compares the calculated (curves *3,4*) and experimental (curves *1,2*) data for a SiC sample implanted at 400 °C. It is evident from the calculations that recombination of the vacancies and interstitials remaining after implantation at 400 °C in the course of annealing at 1200 °C leads to a decrease in the total concentration of defects in the sample (curves *3* and *4* in Fig. 4). During irradiation at 600 °C the vacancies and free interstitials recombine already during the implantation process, so that the curves for the irradiated and annealed samples are virtually identical (Fig. 3).

Near the surface (for $r<50$ nm) the computed curves can differ from the experimental curves. This is due to the fact that the model neglects the impurities arising on the surface when the samples are prepared (see he peak in the experimental distribution of defects, curves *1* and *2* in Figs. 3 and 4).

In the course of the calculations[53,54] and in the present work, certain kinetic parameters used in Eqs. (1)–(7) were varied. The best agreement between the computed and experimental profiles was obtained for the following values of these parameters: migration activation energy of interstitial Si atoms in silicon carbide $\varepsilon_{iS}^{m}\approx 1.55$ eV, recombination parameters in the carbon and silicon subsystems $\mu_C\approx 4\pi\times 10^{-12}$ cm and $\mu_S\approx 4\pi\times 10^{-10}$ cm, respectively; the factors

$$K(T_1)=0.21\,\text{eV} \quad (T_1=400\,^\circ\text{C}) \tag{13a}$$

and

$$K(T_2)=0.49\,\text{eV} \quad (T_2=600\,^\circ\text{C}). \tag{13b}$$

The relations (13a) and (13b) show that the interaction energies of migrating interstitials interacting with the internal stress fields (see Eq. (11)) due to the implanted ions and complexes are different for samples irradiated at different temperatures. This can be explained by the temperature dependence of the defect structures formed on implantation and the temperature dependence of the internal stress fields produced by these structures. For example, the size of the interstitial complexes could increase with implantation temperature; this should be reflected as an increase of the strengths of their stress fields. It is evident from Eqs. (11) and (13) that the interaction energy of interstitials interacting with the stress field for the sample irradiated at 600 °C is greater than for the sample irradiated at 400 °C. In addition, in samples implanted at 800 °C and annealed, defects that can be identified as interstitial dislocation loops[47] form at depths of the order of 100 nm.

CONCLUSIONS

In the present work, taking into account the elastic compression stress fields produced by implanted ions and complexes of interstitial atoms, the model which we proposed in Ref. 53, describing the evolution of defects in silicon carbide samples with high-temperature, high-dose implantation of Al^+ and N^+ ions and also with subsequent annealing, was elaborated. On the basis of this model defect distributions were calculated in satisfactory agreement with the experimental RBS data. Comparing the theoretical and experimental results made it possible to establish a hierarchy of defects formed during implantation. Numerical estimates were obtained for individual kinetic parameters of silicon carbide.

This work was partially supported by Grants No. 97-02-18090 and 96-02-17952 from the Russian Fund for Fundamental Research and Project No. 467-97 from the Moscow Science and Technology Center.

[1)]The sample irradiated at 800 °C is not considered in the model, since transmission electron microscopy (TEM[51]) showed this sample to have a special multilayer structure that is different from the structure of the samples irradiated at other temperatures.

[1] G. L. Harris, *Properties of Silicon Carbide* (INSPEC, London, 1995), 339 pp.
[2] K. Shenai, R. S. Scott, and B. J. Baliga, IEEE Trans. Electron Devices **ED-36**, 1811 (1989).
[3] M. Ruff, H. Mitlehner, and R. Helbig, IEEE Trans. Electron Devices **ED-41**, 1040 (1994).
[4] G. Pensl and Th. Troffer, Solid State Phenom. **47–48**, 115 (1996).
[5] W. J. Choyke and G. Pensl, MRS Bull. , No. 3, 25 (1997).
[6] J. R. O'Connor and J. Smilttens, *Silicon Carbide, a High-Temperature Semiconductor* (Pergamon Press, New York, 1960), 435 pp.
[7] R. F. Davis, G. Kelner, M. Shur, J. W. Palmour, and J. A. Edmond, Proc. IEEE **79**, 677 (1991).
[8] J. A. Powell, P. G. Neudeck, L. G. Matus, and J. B. Petit, Mater. Res. Soc. Symp. Proc. **242**, 495 (1992).
[9] A. Verma and P. Krishna, *Polymorphism and Polytypism in Crystals* (Wiley, New York, 1966), 233 pp.
[10] P. Pirouz and J. W. Yang, Ultramicroscopy **51**, 189 (1993).
[11] V. Heera and W. Skorupa, Mater. Res. Soc. Symp. Proc. **438**, 241 (1997).
[12] J. H. Edgar, J. Mater. Res. **7**, 235 (1992).
[13] W. Wesch, Nucl. Instrum. Methods Phys. Res. **116**, 305 (1996).
[14] G. R. Fisher and P. Barnes, Philos. Mag. B **61**, 217 (1990).
[15] P. Evino, J. Li, A. M. Huntz, and J. Chaumont, Mater. Sci. Eng. B **11**, 331 (1992).
[16] P. Dunbar and P. Birnie, J. Am. Ceram. Soc. **69**, 33 (1986).
[17] A. O. Konstantinov, Fiz. Tekh. Poluprovodn. **26**, 270 (1992) [Sov. Phys. Semiconductors **26**, 151 (1992)].
[18] H. Huang and N. Chonien, J. Nucl. Mater. **212–215**, 148 (1994).
[19] E. N. Mokhov, Yu. A. Vodakov, and G. A. Lomakina, Fiz. Tverd. Tela (St. Petersburg) **11**, 519 (1969) [Phys. Solid State **11**, 415 (1969)].
[20] L. J. Kroto and A. G. Mitnes, Solid-State Electron. **9**, 1125 (1996).
[21] J. A. Spitznagel, S. Wood, and W. J. Choyke, Nucl. Instrum. Methods B **16**, 237 (1986).
[22] R. F. Davis, Thin Solid Films **181**, 1 (1989).
[23] J. A. Edmond, S. P. Withrow, H. S. Kong, and R. F. Davis, Mater. Res. Soc. Symp. Proc. **51**, 395 (1986).
[24] Mc C. J. Hargue, and J. M. Williams, Nucl. Instrum. Methods B **80/81**, 889 (1993).
[25] A. Fohl, R. M. Emrick, and H. D. Carstanjen, Nucl. Instrum. Methods B **65**, 335 (1992).
[26] Mc C. J. Hargue, D. L. Joslin, and J. M. Williams, Nucl. Instrum. Methods B **46**, 185 (1990).
[27] G. Derst, C. Wilbertz, K. L. Bhatia, W. Kratshmer, and S. Kalbitzer, Appl. Phys. Lett. **54**, 1722 (1989).
[28] E. Wendler, A. Heft, and U. Zammit, Nucl. Instrum. Methods Phys. Res. B **116**, 396 (1996).
[29] W. Wesch, A. Heft, and J. Heindl, Nucl. Instrum. Methods B **106**, 339 (1995).
[30] J. Petzold, A. A. Kalnin, D. R. Moskwina, and W. D. Savelyev, Nucl. Instrum. Methods B **80/81**, 8943 (1993).
[31] A. Pérez-Rodríguez, Y. Pacaud, and L. Calvio-Barro, J. Electron. Mater. **25**, 541 (1996).
[32] Y. Pacaud, J. Stoemenos, and G. Brauer, Nucl. Instrum. Methods B **120**, 177 (1996).
[33] Y. Pacaud, W. Skorupa, and J. Stoemenos, Nucl. Instrum. Methods B **120**, 181 (1996).
[34] J. Gardner, M. V. Rao, and O. W. Holland, J. Electron. Mater. **25**, 885 (1996).
[35] E. Glaser, A. Heft, and J. Heindl, Inst. Phys. Conf. Ser. **142**, 557 (1996).
[36] T. Toda, K. Yagi, K. Koga, K. Yoshida, and T. Niina, in *Proceedings of the 6th ICSCRM. IOPC. Ser. 142* (IOP Publ., 1996), pp. 545–551.
[37] S. Ahmed, C. J. Barbero, and T. W. Sigmon, J. Appl. Phys. **77**, 6194 (1995).
[38] M. Behar, P. F. P. Fichtner, P. L. Grande, and F. C. Zawislak, Mater. Sci. Eng. **15**, 1 (1995).
[39] V. Heera, J. Stoemenos, R. Kogler, and W. Skorupa, J. Appl. Phys. **77**, 2999 (1995).
[40] E. Albertazzi and G. Lulli, Nucl. Instrum. Methods B **120**, 147 (1996).
[41] P. G. Neudeck, J. Electron. Mater. **24**, 283 (1995).
[42] B. Culter, P. D. Viber, and W. Rafaniello, Nature **275**, 434 (1978).
[43] W. Rafaniello, K. Cho, and Y. Yrkar, J. Mater. Sci. **16**, 3 (1981).
[44] A. Zangvil and R. Ruh, J. Am. Ceram. Soc. **71**, 884 (1988).
[45] R. S. Kern, S. Tanaka, and R. F. Davis, Inst. Phys. Conf. Ser. **137**, 389 (1994).
[46] R. A. Yankov, N. Hatzopoulos, and W. Fukarek, Mater. Res. Soc. Symp. Proc. **438**, 271 (1997).
[47] J. Pezoldt, R. A. Yankov, W. Fukarek *et al.*, in *Proceedings of the 39th Electronic Materials Conference*, Colorado, USA (1997), to be published in J. Electron. Mater.
[48] R. A. Yankov, M. Fel'son, W. Kreissig *et al.* Pis'ma Zh. Tekh. Fiz. **23**(16), 6 (1997) [Tech. Phys. Lett. **23**, 617 (1997)].
[49] V. S. Kharlamov, D. V. Kulikov, Yu. V. Trushin *et al.*, *Program and Abstracts of International Workshop on New Approaches to Hi-Tech Materials 97 (Nondestructive Testing and Computer Simulations in Materials Science and Engineering)*, St. Petersburg (1997), pp. C2–C13.
[50] Yu. V. Truschin, R. A. Yankov, and V. S. Kharlamov, Mater. Sci. Forum **264–268**, 757 (1998).
[51] R. A. Yankov, W. Fukarek, and M. Voelskow, Mater. Sci. Forum **264–268**, 753 (1998).
[52] Yu. V. Truschin, *Theory of Radiation Processes in Metal Solid Solutions*, (Nova Science Publishers Inc., New York, 1996), 405 pp.
[53] D. V. Kulikov, Yu. V. Trushin, R. A. Yankov *et al.*, Pis'ma Zh. Tekh. Fiz. **24**(1), 39 (1997) [Tech. Phys. Lett. **24**, 17 (1997)].
[54] D. V. Kulikov, J. Pezoldt, P. V. Rybin *et al.*, *Program and Abstracts of International Workshop on New Approaches to Hi-Tech Materials 98 (Nondestructive Testing and Computer Simulations in Materials Science and Engineering)*, St. Petersburg (1998), p. E7.
[55] D. V. Kulikov, R. A. Suris, and Yu. V. Truschin, Supercond. Sci. Technol. **8**, 303 (1995).

Translated by M. E. Alferieff

Sources of acoustoelectric conversion in concretes

T. V. Fursa, N. N. Khorsov, and E. A. Baturin

Tomsk Polytechnical University, 634034 Tomsk, Russia
(Submitted May 19, 1998)
Zh. Tekh. Fiz. **69**, 51–55 (October 1999)

Acoustoelectric conversion in concretes is investigated by physical modeling methods. It is shown by means of equivalent circuit calculations of the electric field in the sample–detector system that the surface of the sample has no effect on the parameters of the signal. An investigation of the temporal and spectral characteristics of acoustoelectric conversion in model and real concretes confirms the calculations. © *1999 American Institute of Physics.* [S1063-7842(99)00710-2]

Concretes and structures made of concrete are most often utilized in places where they are exposed to environmental conditions and are subjected to static and dynamic loads. Under such conditions the probability of unforeseen fracture during service is high. For this reason, the problem of fracture diagnostics is of great practical importance. As was shown in our previous investigations,[1,2] the phenomenon of electromagnetic emission can be used for monitoring fracture processes in concretes and for diagnostics of the strength of concretes.

In our testing method the electromagnetic emission arising in the detection zone upon the dynamic excitation of a sample of concrete is a superposition of the signals determined by the acoustoelectric characteristics of the grains of the aggregate, the contact of the aggregate with the cement base, the cement base itself, the surface of the sample, and the geometry of these elements with respect to the detector.

Concretes usually fracture along the contact of the aggregate and the cement base. In addition, the state of the exterior surface also influences the mechanical characteristics of concretes. To develop an electromagnetic method for monitoring the mechanical characteristics of concretes it is important to determine the sources of electromagnetic emission in the interior volume and on the surface of concrete and to determine the contribution of each source to the total signal. It is especially important to identify the fraction of the electromagnetic response that is directly due to internal defects, because they are what most strongly affects the mechanical characteristics of concrete. Moreover, the depth of a defect zone can be determined from the detection time of the electromagnetic response from an internal defect relative to the time of impact.

Special experiments were performed to investigate the effect of internal sources of electromagnetic emission on the parameters of the electromagnetic signal accompanying the dynamic excitation of concretes. A model sample was prepared so that the electromagnetic response accompanying the excitation of the internal source of electromagnetic emission would be separated in time from that of the source due to vibrations of the sample surface, which is polarized on account of induction from the internal source. A piezoelectric element, from whose faces electric wires were brought to the surface of the sample, was placed into a cement–sand mixture from which a $100\times100\times100$ mm sample was fabricated. The scheme of the experiment was as follows. One surface of the model sample was subjected to a dynamic impact, and two electrical signals were recorded simultaneously using the two-channel ''Emission'' device:[3] one signal directly from the piezoelectric element and the other from an electromagnetic pickup placed 3 mm from a surface oriented perpendicular to the impacted surface. To estimate the travel time of the shock wave along the sample to the surface near which the electromagnetic pickup was placed, the delay time between the electrical signals from two piezoelectric sensors, one located on the sample surface and the other at a definite depth inside the sample, was measured and the propagation velocity of an elastic wave in this sample was determined. In this experiment the piezoelectric element was located 50 mm from the electromagnetic pickup, and the approximate arrival time of the acoustic wave at the surface near which the electromagnetic pickup was located was 50 μs. At an instrumental resolution ≈1 μs, time delays could be easily determined for spatial separations of the signal sources as small as 1 mm. If the electromagnetic response is detected directly from an interior source (a piezoelectric element), the delay time of this signal compared with the detection of a signal along the wires is determined by the polarization time of the sample ($<10^{-10}$ s), i.e., there is essentially no delay (Fig. 1). It is seen from the figure that these two signals appear at the same time. When the electromagnetic pickup was replaced by an acoustic pickup attached to the same surface that the electromagnetic pickup had been mounted near, it was found that the arrival time of the electromagnetic (a) and acoustic (b) signals was shifted by 50 μs (Fig. 2).

If the piezoelectric element is replaced by gravel in the cement–sand mixture, the picture of the process remains the same, i.e., if electromagnetic and acoustic signals are simultaneously detected from the same surface of the sample, then the time delay in the appearance of these pulses is the same 50 μs, provided that the gravel is located inside the sample at the same depth as the piezoelectric element.

The temporal character of the signals detected with an electromagnetic pickup is very close to that of the piezoelec-

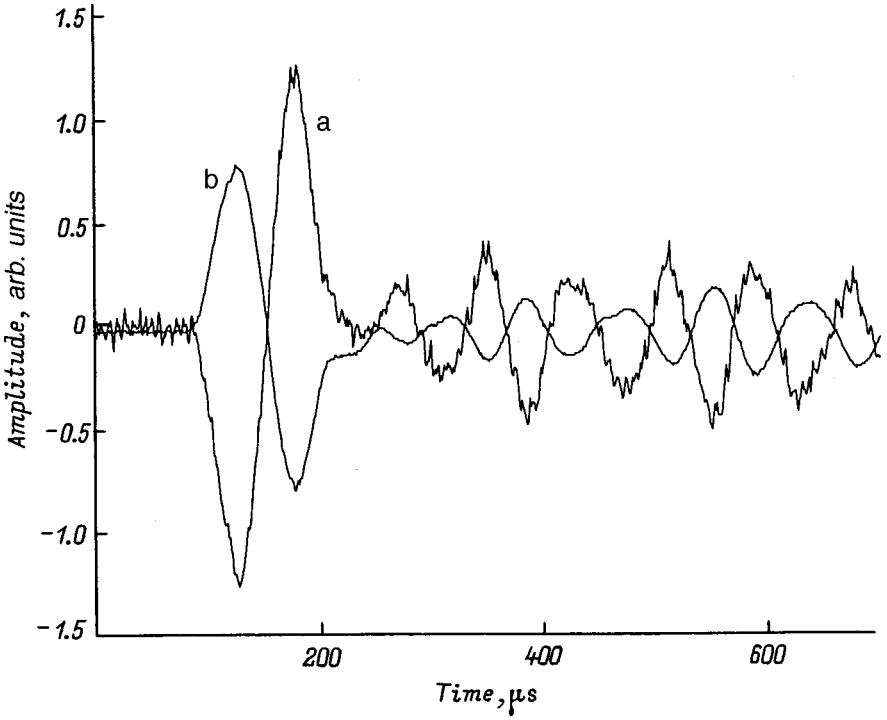

FIG. 1. Electrical response (a) and electrical signal from a piezoelectric element (b) which are detected as a result of impact excitation of the model sample.

tric element. To determine whether or not there are differences in the signals, their spectral characteristics were calculated. Comparing the spectral energy density of the pulses from the piezoelectric element and the electromagnetic pickup (Fig. 3) shows that the spectrum of the electromagnetic signal (a) is essentially identical to that of the piezoelectric element (b); the differences are negligible. Comparing the spectral energy density of the signals detected from the piezoelectric element (Fig. 3b) and the acoustic pickup (Fig. 3c) shows that the spectra of these pulses are substantially different. Specifically, components below 5 kHz, which are present in the acoustic pulse, are virtually absent

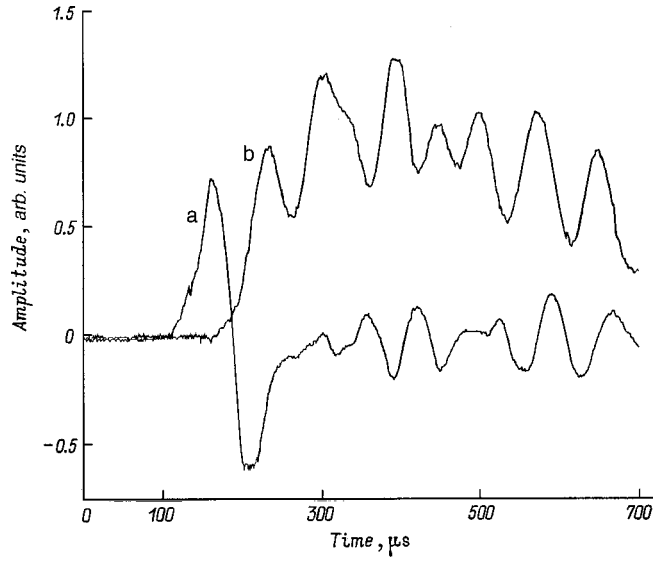

FIG. 2. Acoustic oscillations of the surface (a) and electric signal from a piezoelectric element (b), which are detected as a result of dynamic excitation of the model sample.

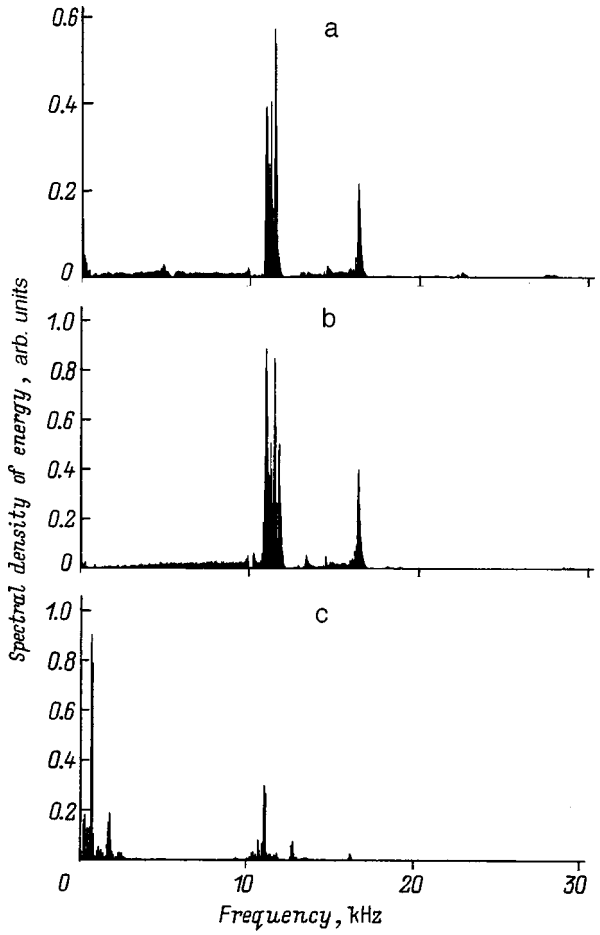

FIG. 3. Spectral characteristics of pulses detected as a result of dynamic excitation of the model sample: a — electromagnetic response, b — signal from the internal piezoelectric element, c — acoustic signal.

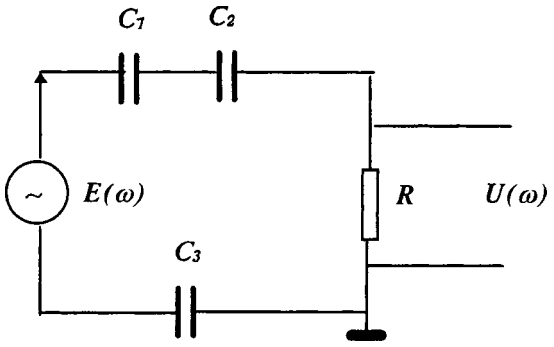

FIG. 4. Equivalent circuit for the acoustic-electric conversion signal.

in the spectrum of the signal from the piezoelectric element; these components are the ones that could be caused by surface waves, near which the electromagnetic pickup is located.

To estimate the contribution of surface oscillations to the resulting electromagnetic signal, detected with impact excitation of the experimental object, theoretical calculations of this contribution were performed using the measurement scheme indicated above. A capacitive pickup senses a displacement current, whose sources are acoustoelectric conversions. The displacement current is modeled by oscillations of the sample surface turned toward the detecting face of the pickup. The magnitude of the current is determined by the distributed and lumped parameters of a closed electric circuit and can be described by the equation for the field of an electric current[4]

$$\nabla \cdot \left(\gamma \mathbf{E} + \varepsilon \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) = 0, \tag{1}$$

where $\mathbf{E}$ is the intensity of the electric field driving the motion of the electric charges, γ is the conductivity of the electric circuit, and ε and ε_0 are, respectively, the dielectric constant (relative permittivity) of the medium carrying the displacement current and the permittivity of free space, respectively; ∇ is the differential operator

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}.$$

To assess the effect of surface oscillations on the magnitude of the current (the percent of modulation) Eq. (1) must be solved (for example, numerically) with allowance for the temporal factors of the change in the boundary conditions. At the same time, it is easier to estimate the percent modulation, replacing the above-described scheme with distributed parameters by an equivalent electric circuit with lumped parameters (Fig. 4).

In the circuit $E(\omega)$ is the emf produced in a part of the circuit; C_1 is the capacitance of the system comprising the acoustoelectric transducer plus the sample surface facing toward the signal detector; C_2 is the capacitance of the system comprising the sample surface plus the detector face; C_3 is the capacitance of the system comprising the acoustoelectric transducer plus the grounded surface that closes the circuit and includes the internal impedance of the emf part of the circuit; R is the input resistance of the capacitive pickup (we neglect the input capacitance C_{in} at angular frequencies $\omega < 1/RC_{\text{ent}}$); $U(\omega)$ is the output voltage. The output voltage is determined by the formula

$$U(\omega) = E(\omega) \cdot \frac{R}{(1/j\omega C_1 + 1/j\omega C_2 + 1 j\omega C_3 + R)}. \tag{2}$$

The capacitances C_1, C_2, and C_3 can be represented in the form

$$C_1 = \varepsilon \cdot \varepsilon_0 \frac{S}{d_1}, \quad C_1 = \varepsilon_0 \frac{S}{d_2}, \quad C_3 = \varepsilon \cdot \varepsilon_0 \frac{S_3}{d_3},$$

where S is the effective area of the two-layer capacitor, S_3 is the effective area of the closing capacitance, d_1 is the effective distance from the signal source to the sample surface, d_2 is the effective distance from the surface to the face of the capacitive pickup, d_3 is the effective distance from the current source to the grounding surface; ε is the dielectric constant of the sample, and ε_0 is the permittivity of free space.

It should be noted that the representation of sections of an electric circuit where a displacement current flows by flat capacitors is arbitrary. The justification for such a representation that the dependence of the change in the capacitance on the distance between the plates is stronger for a flat capacitor than for spherical or cylindrical capacitors, i.e., the most favorable situation with respect to modulation of the signal by oscillating surfaces obtains.

Substituting the values of the capacitances into Eq. (2) and performing a transformation, we obtain

$$U(\omega) = E(\omega) R j\omega \varepsilon \varepsilon_0 S \frac{1}{(d_1 + \varepsilon d_2) + d_3 S/S_3 + j\omega \varepsilon \varepsilon_0 S R}. \tag{3}$$

The increment ΔU to the output voltage due to the change in the distance d_1 by an amount δd and d_2 by an amount $-\delta d$ (d_1 and d_2 change in antiphase) can be found by representing Δd as a differential of the function $U = f(d_1, d_2)$ (here and below the argument ω in U and E is dropped for simplicity),

$$\Delta U = \frac{\partial U}{\partial d_1} \delta d + \frac{\partial U}{\partial d_2}(-\delta d). \tag{4}$$

Substituting expression (3) into Eq. (4) we obtain

$$\Delta U = E \cdot R \cdot j\omega \cdot \varepsilon \cdot \varepsilon_0 \cdot S \times \frac{\delta d \cdot (\varepsilon - 1)}{(d_1 + \varepsilon \cdot d_2 + d_3 S/S_3 + j\omega \cdot \varepsilon \cdot \varepsilon_0 \cdot S \cdot R)^2}. \tag{5}$$

We define the modulation factor K as

$$K = \frac{\Delta U}{U} = \frac{\delta d \cdot (\varepsilon - 1)}{d_1 + \varepsilon \cdot d_2 + d_3 S/S_3 + j\omega \cdot \varepsilon \cdot \varepsilon_0 \cdot S \cdot R}. \tag{6}$$

Approximate calculations taking into account the energy of the impact, the duration of the impact (the width of the excitation spectrum), the spectral components of the surface waves, and the elastic constants of concrete give for δd a value of the order of 10 microns for the most favorable conditions, while d_1 and d_2 can vary from several millimeters to

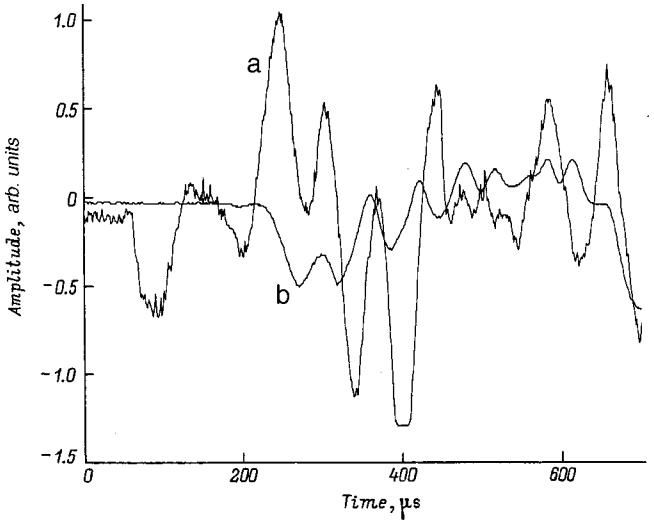

FIG. 5. Electromagnetic response (a) and acoustic oscillation of the surface (b) which are detected as a result of dynamic excitation of a concrete beam.

tens of centimeters. Thus, as one can see from the last expression, modulation essentially does not contribute to the detected signal.

Therefore, when a source of electromagnetic emission is present inside the sample, the parameters of the electromagnetic response from the sample into which this source is placed are determined virtually completely by its own acoustoelectric characteristics. The delay time in the appearance of the electromagnetic response compared with the time of impact can serve as a criterion for determining the depth of the internal source of the acoustoelectric transformations (a defect) inside the sample relative to the point of impact. However, if the sample contains a large number of sources, then the minimum delay time of the electromagnetic response will determine the depth of the source closest to the impacted surface. To determine the depth of the most dangerous defect and the parameters of this defect, further investigations in this direction must be performed.

A picture similar to that presented in the present model experiments is observed in tests of concrete with a random distribution of sources of electromagnetic emission and the source of electromagnetic emission in the form of a grain of the aggregate that emerges at the surface of the concrete is subjected to dynamic excitation. As an example, we present the results of the following experiment. A $1200\times130\times65$ cm concrete beam was used for the investigations. One grain of the aggregate emerging at the surface was mechanically excited with an impactor. A beam was used so as to be able to detect electromagnetic and acoustic signals when they are separated by quite large distances. At first, the electromagnetic and acoustic pickups were placed at the same distance (25 cm) from the impactor (Fig. 5). It is evident from the figure that at first an electromagnetic emission signal is detected, after which the acoustic signal appears with a 165 μs delay. If the electromagnetic pickup is left at its previous location and the acoustic pickup is moved in the direction of the impactor, then the delay time decreases in proportion to the change in the distance. The minimum delay is recorded when the acoustic pickup is located at the shortest possible distance from the impactor. Once again the delay increases when the acoustic pickup is moved in a direction opposite to the electromagnetic pickup and the impactor.

In addition, the electromagnetic response for tests performed on beams with a more complicated structure, than the model sample consisting of a cement base and one piezoelectric element, because of the presence of a large number of grains of the aggregate with nonuniform mechanical and electrical properties, is more complicated (Fig. 5a). Evidently, in this case the electromagnetic response from many sources lying in the propagation path of the acoustic wave through the beam is detected.

This example presents a concrete case of excitation of a source of electric signals (impact on the aggregate emerging at the surface of the beam). However, in complicated heterogeneous systems such as concretes the acoustoelectric conversion picture is ordinarily more complicated. If the impact is made at different locations on the surface of the beam, then depending on the arrangement of the internal sources of the acoustoelectric transformations, which the grain of the aggregate are, and their elastic and electric characteristics, different variants, with respect to the time of appearance of acoustic and electromagnetic signals, are observed. Cases where acoustic and electromagnetic pickups are located at the same distances from the impactor and the signals from these pickups are detected simultaneously are sometimes observed. This is due to the fact that an acoustic detector is accidentally located near a source of acoustoelectric conversion, whose electric field energy is higher than the detection threshold of the detector. Because the acoustoelectric conversion factor of other sources, which an elastic wave can reach earlier, is small, these waves are not detected by the electromagnetic pickup, i.e., the acoustoelectric conversions occurring in complex heterogeneous systems such as concretes are also very complicated and require further investigations.

These investigations show that the grains of the aggregate play a determining role in the generation of electromagnetic emission of concretes under dynamic excitation. Oscillations of the sample surface due to mechanical excitation have virtually no effect on the acoustoelectric conversion signal. The depth of an internal defect can be determined, in principle, from the arrival time of electromagnetic pulses relative to the time of excitation.

[1] V. L. Chakhlov, Yu. P. Malyshkov, V. F. Gordeev *et al.*, Izv. Vyssh. Uchebn. Zaved. Stroitel'stvo, No. 5–6, 54 (1995).
[2] Yu. P. Malyshkov, V. F. Gordeev, T. V. Fursa *et al.*, Izv. Vyssh. Uchebn. Zaved. Stroitel'stvo, No. 5, 3 (1995).
[3] V. F. Gordeev, V. P. Eliseev, Yu. P. Malyshkov *et al.*, Defektoskopiya, No. 4, 48 (1994).
[4] V. A. Govorkov, *Electric and Magnetic Fields* (Énergiya, Moscow, 1968).

Translated by M. E. Alferieff

Deformation autowaves in nitrogen-doped γ-Fe single crystals

L. B. Zuev, S. A. Barannikova, and V. I. Danilov

Institute of Physics of Strength and Materials Engineering, Siberian Branch of the Russian Academy of Sciences, 634021 Tomsk, Russia

Yu. I. Chumlyakov and I. V. Kireeva

Siberian Physicotechnical Institute at Tomsk State University, 634050 Tomsk, Russia
(Submitted May 22, 1998; resubmitted April 7, 1999)
Zh. Tekh. Fiz. **69**, 56–62 (October 1999)

Experimental data on the evolution of macrodeformation fields of single-crystal samples of austenitic chromium–nickel steel with a superequilibrium nitrogen content under tension are analyzed on the basis of a model of autowave plastic flow. The conditions for the appearance and observation of different types of autowave deformation structures are established; such structures include a solitary front, a traveling autowave, and a stationary dissipative structure and are determined by the crystal-geometry of deformation and by the nitrogen concentration in the material. It is shown that a one-to-one correspondence exists between the type of deformation autowave and the stage of the plastic flow curve of the material. [S1063-7842(99)00810-7]

INTRODUCTION

The stability of the plastic flow of materials during pressure treatment, the longevity, and the fracture of parts and structures under a load depend strongly on the degree of macroscopic uniformity of the deformation. Although the facts concerning its inhomogeneity on this scale have been known for a long time,[1–3] usually it has been assumed that this effect is immaterial and the distribution of local deformations over the volume of the loaded object can be assumed to be quasiuniform at least up to the high degrees of deformation, for which a neck forms. The works of the author of Ref. 4 are an exception. He showed that the foci of plastic deformation of a sample under active loading and creep move in a regular manner, creating a so-called ''traveling'' neck.

Fundamentally new results, making it possible to determine the quantitative parameters of the evolution of the distributions of local deformations, have been obtained using the speckle-interferometry method to investigate the inhomogeneity of deformation.[5–8] It was found that in the course of plastic deformation three types of spatiotemporal distributions of local defects can be distinguished, and there exists a unique correspondence between the law of plastic flow $\sigma(\varepsilon)$ and the type of spatiotemporal evolution of the deformation distribution. The following can be stated on the basis of an analysis of the experimental data of Refs. 5–8: i) on the ''yield bench'' (the section of easy slip), where the work hardening coefficient $\Theta = d\sigma/d\varepsilon \approx 0$, an isolated front of plastic deformation moves through the sample; ii) at the linear hardening stage ($\sigma = \Theta\varepsilon$, $\Theta = \text{const}$) an equidistant sequence of localized deformation zones (autowaves) moves with a constant velocity through the sample; iii) at the parabolic hardening stage ($\sigma \sim \varepsilon^{1/2}$) a stationary pattern of localized deformation foci is formed; iv) when the law of work hardening changes, the type of localization of deformations changes. At the transitional stage a chaotic distribution of the components of the plastic distortion tensor is observed.

These laws are quite universal, since they were established in experiments performed on a representative range of materials in single- and polycrystalline states, single- and multiphase states, with face-centered crystal (fcc), body centered crystal (bcc), and hexagonal close-packed (hcp) lattices. Their generality has made it possible to formulate the hypothesis that this process is of an autowave character.[9] However, to refine the ideas about the nature of autowaves it is necessary to determine whether their forms are associated with large differences in the kinetics of plastic flow of materials with different structure or they can also be observed in a single substance whose work hardening varies, for example, as a result of additional doping, which does not change the structure of the alloy but influences its plasticity.

MATERIALS AND METHOD OF THE INVESTIGATIONS

Single crystals of additionally nitrogen-doped austenitic steel with the composition (in mass %) C (0.013), Cr (18.0), Ni (12.4), Mo (2.3), Mn (1.2), V (0.01), and Si (0.06) are convenient for this purpose. The parameters of work hardening are controlled in this case by choosing the crystallographic orientation of the direction of application of the load and the nitrogen content in the γ solid solution.

In the present work single crystals grown by the Bridgman method on a Redmet-1 apparatus were used.[10] After the single-crystal ingot obtained was homogenized for 50 h at 1473 K, samples in the form of a double paddle were cut out on an electric erosion stand in a manner so that their wide face would coincide with the (011) plane and the longitudinal axis would coincide with the [001] or [111] directions. The dimensions of the working part were $25 \times 5 \times 1.6$ mm

TABLE I. The elements of slip and the growth stages in the experimental samples (the dominant slip systems are indicated in boldface).

Nitrogen content, mass %	Orientation of axis of extension	Acting slip system	Strain intervals, %	Stage of the deformation curve
0.35	[001]	$[\mathbf{0\bar{1}1}]$ $(\mathbf{111})$	0–3.3	Elasticity+transitional section;
		$[\mathbf{0\bar{1}1}]$ $(\bar{1}\mathbf{11})$	3.3–7.0	linear stage ($\sigma=\Theta\varepsilon$, $\Theta=1100$ MPa);
		$[\bar{1}01]$ (111)		
		$[\bar{1}01]$ (111)	>7.0	parabolic stage, ($\sigma\sim\varepsilon^n$)
0.35	[111]	[011] $(1\bar{1}1)$	0–1.0	elasticity+transitional section;
		$[\mathbf{101}]$ $(\mathbf{11}\bar{\mathbf{1}})$	1.0–4.0	elasticity tooth+easy slip ($\Theta=250$ MPa);
		$[\bar{1}1\bar{1}]$ $(\bar{1}11)$	4.0–8.0	transitional section;
			8.0–17.0	linear stage ($\sigma=\Theta\varepsilon$, $\Theta=1200$ MPa);
			>17.0	parabolic stage ($\sigma\sim\varepsilon^n$)
0.50	[001]	$[\mathbf{0\bar{1}1}]$ $(\mathbf{111})$	0–1.8	Elasticity+transitional section;
		$[0\bar{1}1]$ $(\bar{1}11)$	1.8–4.0	linear stage ($\sigma=\Theta\varepsilon$, $\Theta=4200$ MPa);
		$[1\mathbf{01}]$ $(\bar{1}\mathbf{11})$	>4.0	parabolic stage ($\sigma=\varepsilon^n$)
0.50	[111]	[011] $(1\bar{1}1)$	0–1.5	Elasticity+transitional section;
		$[\mathbf{101}]$ $(\mathbf{11}\bar{\mathbf{1}})$	1.5–4.0	linear stage ($\sigma=\Theta\varepsilon$, $\Theta=1500$ MPa);
		$[\bar{1}01]$ $(\bar{1}11)$		
			4.5–17.0	linear stage ($\sigma=\Theta\varepsilon$, $\Theta=2000$ MPa);
			>17.0	parabolic stage ($\sigma\sim\varepsilon^n$)

for the [001] samples and $30\times5\times1.6$ mm for the [111] samples. Nitriding of the samples was performed by a gas-temperature-pressure method[11] at 1473 K temperature and saturation atmospheric pressure 1.5×10^5 Pa, which made it possible to obtain the required nitrogen concentration in the solid solution. After nitriding, the samples were once again homogenized in vacuum at 1473 K for 1 h and quenched in water. The final nitrogen content in the experimental samples was 0.35 and 0.5%. This choice of concentration range is based on the fact that according to Ref. 10, in tests of samples with the axis of extension oriented in the [001] and [111] directions at temperature ~300 K and nitrogen concentration less than 0.3 mass %, its influence is negligible and only a small increase of the critical cleavage stress with no change in the course of the work hardening is observed. For 0.35–0.4 mass % nitrogen and the same deformation temperature, together with a substantial increase in the critical cleavage stress, a change was noted in the form of the plastic flow curve of [111] crystals. Doping up to 0.5 mass % N strongly influences the form of the curves $\sigma(\varepsilon)$ of single crystals of both orientations.

The crystal geometry of plastic displacements for the indicated orientations is such that multiple slip is possible in samples of both types immediately after the yield point is reached. For [001] orientation the four slip systems [011] (111), [101] (111), [101] (111), and [011] (111) have a Shmid factor $m=0.41$. In the observation plane (011) they give slip traces at angles of 90° (first two) and 35°16′ (second two) with respect to the axi of extentions. For [111] orientation three systems have the same Shmid factor $m=0.27$: [101] (111), [011] (111), and [$\bar{1}$11] (111). The slip traces from them on the working surface of the sample make angles of 19°28′, 35°16′, and 90°, respectively, with the axis of extension.

The basic experimental procedure consisted in successive reconstruction of the fields of the displacements vectors of points of the sample surface by the method of double-exposure speckle interferometry[12] directly during extension at a rate of 5.5×10^{-5} s^{-1} in an Instron-1185 testing machine. The vector displacement field $\mathbf{r}(x,y)$ was recorded discretely with 30 s intervals, which corresponded to the increment to the total deformation $\delta\varepsilon_{tot}=0.2\%$ (x,y — coordinates of the points on the working surface of the sample). The values of the components of the plastic distortion tensor $\beta=\nabla\mathbf{r}(x,y)$ (local elongation ε_{xx}, shear ε_{xy}, and rotation ω_z)[13] were determined by numerical differentiation of the $\mathbf{r}(x,y)$ data file with respect to the coordinates x and y at points of the sample surface with a 1 mm step. The distributions of the slip traces in the deformed samples were analyzed in a Neophot-21 metallographic microscope. The tests were performed at 300 K.

EXPERIMENTAL RESULTS

The experimental conditions and the data obtained by tensile testing of samples with two orientations of the axis of extension and different nitrogen concentrations are summarized in Table I. In these four cases a complete set of possible types of dependences $\sigma(\varepsilon)$ was observed in different strain ranges. The strain curves for these four variants of the samples are presented in Fig. 1.

The experimental results will be presented below in the-

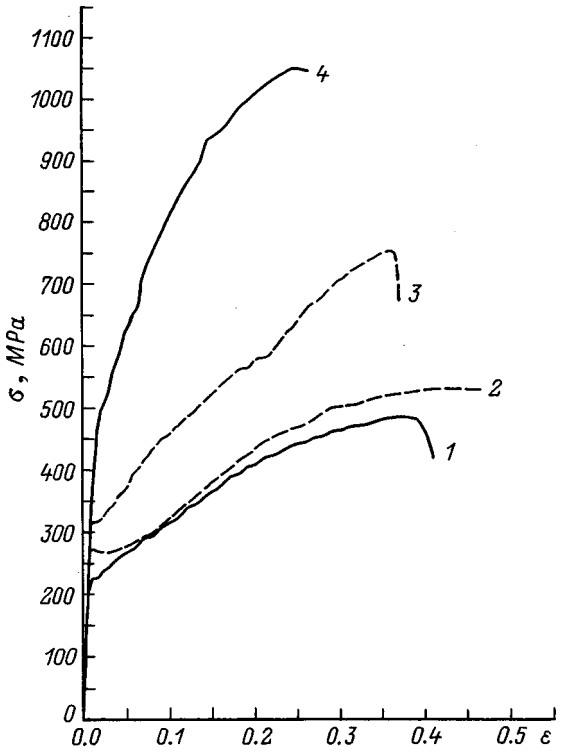

FIG. 1. Typical deformation curves of single crystals of austenitic steel. Nitrogen content (C_N, mass %) and crystallographic orientation of the axis of extension: *1* — 0.35 [001]; *2* — 0.35 [111]; *3* — 0.5 [111]; *4* — 0.5 [001].

form of a description of the spatiotemporal patterns of distribution and motion of the deformation foci in a deformed sample at distinct stages of the plastic flow curve, identified above as i, ii, and iii.

1) *The stage of easy slip* was observed in 0.35 mass % N samples whose axis of extension was oriented along [111] (see Table I). In this case the strain diagrams of the samples (Fig. 1, curve *2*) showed a distinct tooth and a yield bench (section of easy slip), whose overall length is limited to the range $1 \leqslant \varepsilon_{tot} \leqslant 4.5\%$. The existence of the first stage and the predominance of the slip system [101] (111) at this stage are attributed[10] to the instability and localization of the displacement in one of the three equivalent slip systems because of an increase in the splitting of the gliding dislocations with Burgers vectors $a/2\langle 110\rangle$ into partial Schockley dislocations with Burgers vectors $a/6\langle 110\rangle$ as a result of a decrease of the effective energy of a stacking fault under the action of an external tensile stress.[10,14]

The distributions of local elongations ε_{xx} in the process of deformation of these crystals was analyzed in the range $1.2\% \leqslant \varepsilon_{tot} \leqslant 11\%$, covering the section of easy slip, the transitional stage, and part of the linear stage. Up to $\varepsilon_{tot} = 2\%$, no stable regularities were found in the arrangement of the zones of localization of ε_{xx} relative to one another and the pattern was completely chaotic. Then a spatial pattern consisting of three equidistant maxima of local elongations, whose dynamics (Fig. 2) was characterized by the following, formed against the chaotic background. Of the three foci arising initially, one moved with velocity $\sim 3.5 \times 10^{-5}$ m/s over the entire extent of the easy-slip stage. The two others soon stopped and remained stationary up to the end of the stage. Thus, motion of a solitary plasticity front was ob-

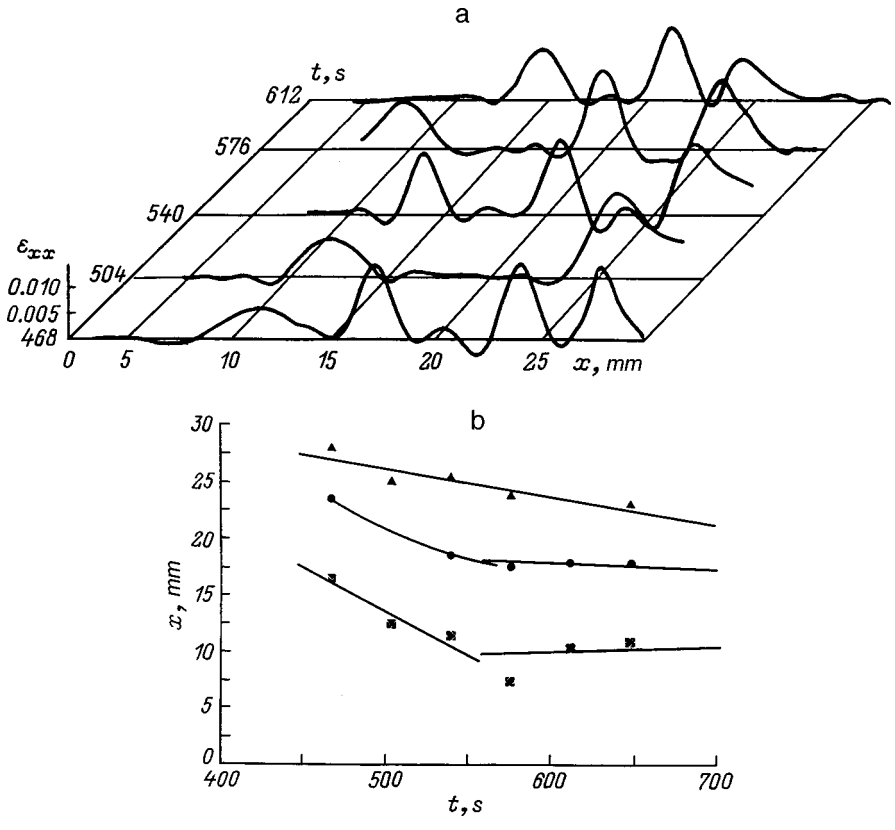

FIG. 2. Evolution of the distributions of local elongations ε_{xx} at the easy-slip stage of single crystals of high-nitrogen steel (a) and dependence of the positions of the maxima of ε_{xx} on the loading time (b). $C_N = 0.35$ mass %, [111] axis of extension. Symbols — positions of the deformation maxima at different times.

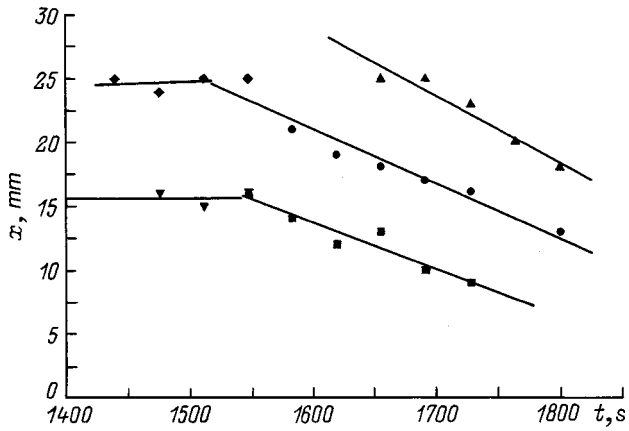

FIG. 3. Positions of local maxima ε_{xx} as a function of the loading time at the stage of linear hardening of single crystals of high-nitrogen steel with [111] axis of extension and $C_N = 0.35$ mass %. Symbols — same as in Fig. 2b.

served at the easy-slip stage. At the end of this stage, for $\varepsilon_{tot} > 4\%$, the distributions of local elongations, just as other components of the distortion tensor, once again became chaotic.

2) *Linear hardening* ($\sigma \sim \varepsilon$, $\Theta = \text{const}$) commenced with further extension of this single crystal ($\varepsilon_{tot} > 8\%$) and was accompanied by the creation of three moving equidistant deformation zones from the chaos. This is shown in Fig. 3 in the form of a dependence of the x coordinates of the maxima of ε_{xx} on the deformation time. One can see that synchronous motion of all three foci commences with a certain delay, but after the regime is established, the characteristic distance between them $\lambda \approx 7 \pm 1$ mm remains throughout the extension process. The direction of motion of the foci is the same as at the easy-slip stage, and the velocity, calculated from the slope of the straight lines, is $v \approx 5.0 \times 10^{-5}$ m/s. The position of such foci on the frontal plane of the sample was normal to the axis of extension.

A similar picture was observed in the tensile testing of a single crystal with 0.35 mass % N and [001] orientation. In this case it was possible to record the process establishing self-consistent velocities of the foci of plastic flow (Fig. 4),

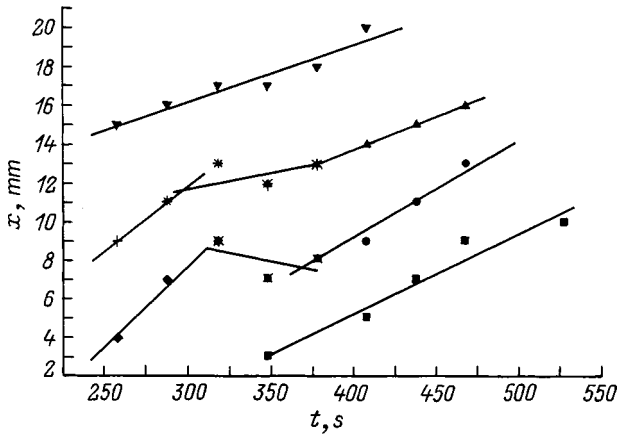

FIG. 5. Same as Fig. 3 but with [001] axis of extension and $C_N = 0.5$ mass %. Symbols — same as in Fig. 2b.

occuring in a manner such that the foci of localization of deformation appearing later increase in velocity, so that ultimately an autowave structure is forms.

The stages of linear hardening were also observed in austenite nitrided up to 0.5 mass % N for both orientations of the axis of extension (see Table I). In the case of extension along [001], starting at $\varepsilon_{tot} \approx 2\%$ a system of four equally spaced maxima of ε_{xx} with spatial period $\lambda = 4.5 \pm 1$ mm formed (Fig. 5). Here, at first an adjustment of the velocities is observed also, after which this group of maxima moved during the entire linear stage up to $\varepsilon_{tot} = 4\%$ with virtually the same velocities $\sim 2.6 \times 10^{-5}$ m/s, forming a typical autowave pattern.

An interesting feature was observed in an investigation of the plastic flow curve of a single crystal containing 0.5 mass % N with a [111] axis of extension. In this case (Fig. 6) the two foci of deformation formed at the beginning of the linear hardening stage and moving synchronously with the same velocity stopped successively at the same location in the sample (approximately 10 mm from the end), after which their motion resumed with the same velocity. We note that the second focus stops for a 1.5 times shorter time than first one. Next, at the same stage but with larger deformations, motion of a similar bound pair of such foci is once again

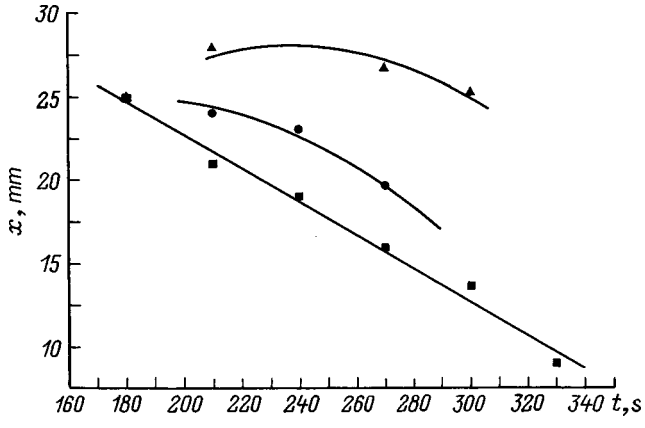

FIG. 4. Same as Fig. 3 but with a [001] axis of extension. Symbols — same as in Fig. 2b.

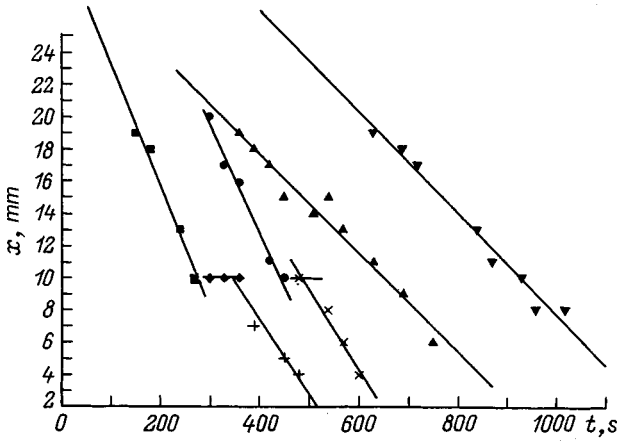

FIG. 6. Same as Fig. 3 but $C_N = 0.5$ mass %. Symbols — same as in Fig. 2b.

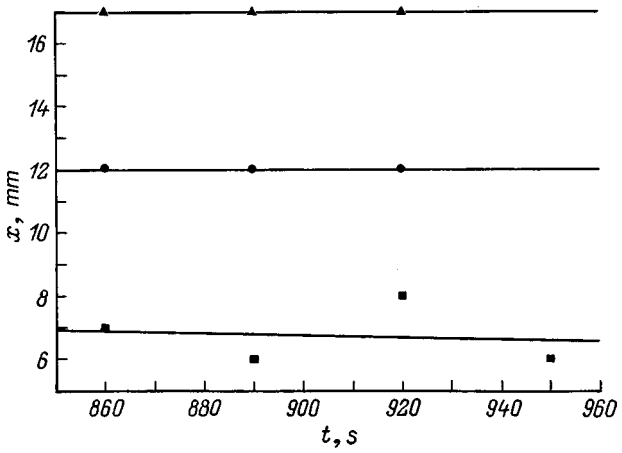

FIG. 7. Positions of the local maxima of ε_{xx} versus the loading time at the parabolic stage of the hardening curve of single crystals of high-nitrogen steel for a [001] axis of extension and $C_N = 0.5$ mass %. Symbols — same as in Fig. 2b.

observed, but this time with a lower velocity (Fig. 6) and no stoppages. The latter, as microscopic investigation of the sample showed, were related with the boundary of a crystallite with a different orientation, randomly formed at the time the sample was grown. Under the indicated conditions, the deformation zones were not perpendicular to the sample axis, but rather they made an angle of approximately 20° with it. This seems to be determined by the fact that the traces from the predominant slip system in this case intersect the observation plane, as indicated above, at an angle of 19°28′ with respect to the axis of extension. A similar crystallographic causation of the arrangement of moving deformation zones had been observed previously in similar investigations of the plasticity of single-phase single crystals Cu + 10%N + 6%Sn.[6] Thus, the formation of an autowave structure is a common feature of the linear hardening stages in all doped austenite single crystals investigated.

3) *Parabolic hardening* ($\sigma \sim \varepsilon^{1/2}$) in the experimental samples completed the deformation process in all four cases investigated (see Table I), but for different values of the total strain (Fig. 1). At these stages a stationary dissipative structure was always observed, i.e., three or four stationary deformation foci were located along the sample with interval 4–7 mm. This pattern is shown in Fig. 7 for the tensile testing of a sample with 0.5 mass % N for a [001] axis of extension.

AUTOWAVE REGIMES OF PLASTIC FLOW

The results obtained, on the one hand, confirm the previously established regularities in the patterns of deformation localization, while in the other they enrich with details our ideas about the dynamics of the development of this phenomenon. For plastic deformation of single crystals of highly doped γ-Fe, all patterns of deformation localization observed previously in different materials were found: solitary front, traveling autowave, and stationary dissipative structure. Additional, to that found previously,[15–17] confirmation was found for the correspondence between the stage structure of the work hardening and the type of patterns of deformation localization which are observed, as described in the introduction. The fact that the same materials were used in the experiments makes it possible to assert that the changes in the patterns of deformation localization are actually controlled only by the nitrogen content in steel and by the orientation of the axis of extension of the samples. This means that a causal link exists between the observed macroscopic patterns of deformation localization and the microscopic mechanisms of elementary acts of plastic flow which are specific to each stage of the process. The two most important common indicators of the processes investigated are the appearance of a macroscopic scale $\lambda \approx 4-7$ nm in the arrangement of deformation foci throughout the sample and the low velocity of the foci $\sim(2-8)\times 10^{-5}$ m/s. At the same time, the elementary acts of deformation of dislocation nature, naturally, do not have such a scale, and the dislocation velocity characteristic for them is ordinarily much higher. The spatial scale of the distributions of local deformations changes little with a transition from one stage of plastic flow to another and with a change in the general form of the deformation curve as a function of the nitrogen content, but it is largely determined by the orientation of the axis of extension. The velocities of deformation foci were of the same order of magnitude in all experiments, irrespective of the nitrogen content and the orientation of the samples. However, it is observed that they tend to decrease as the total deformation of the sample increases during the testing process.

It is evidently impossible to describe these complicated regularities of plastic flow on the basis of the traditional models of the mechanics of a continuous medium. The dislocation theory is also found to be inapplicable for these purposes, primarily because of the large difference of the characteristic spatial parameters: the Burgers vector of a dislocation $\mathbf{b} \approx 10^{-10}$ m and the scale of the observed nonuniformity of plastic flow $\lambda \approx 10^{-2}$ mm $\gg \mathbf{b}$. The observed regularities should be attributed to spatiotemporal self-organization of the deformation field of the sample. Just as in Ref. 18, here self-organization is taken to mean the acquisition of spatial, temporal, or functional structure by the system without any specific external action. Self-organization in systems of different kinds can be successfully described phenomenologically on the basis of the theory of nonequilibrium processes in open systems (synergetics).[18] Attempts in the literature (see, for example, Refs. 19–22) to use such an approach for the problem of plasticity have largely been based on the use of dislocation models (except for Ref. 19, where the evolution of the fields of deformation gradients was studied). In our works,[7,15–17] such laws of localization of deformation have been discussed on the basis of one of the variants of synergetics — the theory of autowave processes.[23] In so doing, plastic deformation and elastic stress were used as variables to characterize the state of the system, since the experimental technique employed makes it possible to fix directly at least the deformation field. In accordance with Ref. 24, these quantities are the autocatalytic (ε) and damping (σ) factors, which control the kinetics of plastic deformation. The differential equations describing their evolution have the form[7,15–17]

$$\varepsilon_t = \chi(\varepsilon,\sigma) + D^{\varepsilon}\varepsilon_{xx}, \tag{1}$$

$$\sigma_t = \phi(\sigma,\varepsilon) + D^{\sigma}\sigma_{xx} \tag{2}$$

(the lower indices indicate differentiation with respect to the corresponding variable).

The first terms on the right-hand sides are related with the continuous (''hydrodynamic'') parts of the strain and stress fluxes, respectively, and the functions $\chi(\varepsilon,\sigma)$ and $\phi(\sigma,\varepsilon)$ are the nonlinear rates of change of the stresses and strains in local volumes and describe the kinetics of elementary acts of plastic deformation[7,17,25] of a dislocation nature. The terms containing the second spatial derivatives of σ and ε give the stochastic (''diffusion'') redistribution of these quantities in space. The quantities D^{ε} and D^{σ} with dimensions of the diffusion coefficient (m^2/s) contain information about the relation between the micro-, meso-, and macroscales of the levels of plastic flow.[16,17,25,26]

Equations (1) and (2) are very difficult to solve. Although a solution has not yet been found, these equations can still be used for qualitative analysis of the character of autowave processes. The parameter controlling the evolution of a self-organization process on the basis of the idea that it is of an autowave nature is the so-called refractory time Θ.[23] For spatiotemporal patterns of deformation localization which are observed at the corresponding stages of plastic flow, it is determined by the dynamics of local elementary acts of displacement at the corresponding stages of the curve $\sigma(\varepsilon)$ and is the time period during which the system remains locally unsusceptible to external actions after a single realization of an elementary displacement. We shall examine below three variants of an assessment of the situation, assuming that at the yield bench stage (easy slip) $\Theta \to \infty$, at the linear hardening stage $\infty > \Theta > 0$, and at the parabolic hardening stage $\Theta \to 0$. These assumptions are physically adequately grounded. Thus, in the case of the yield bench, on which plastic flow occurs by propagation of a Lüders band, on the front of the latter each elementary displacement occurs only once and then the deformation moves to a different location. At the parabolic hardening stage the process is controlled by transverse slip processes, i.e., displacement can be realized at any time. At the linear hardening stage, evidently, a more complicated intermediate situation occurs. Using for D^{ε} the obvious relation[26]

$$D^{\varepsilon} \approx V_f^2 \Theta, \tag{3}$$

where V_f is the velocity of the plastic deformation front, it can be assumed, on the basis of what has been said above for the yield bench, that $D^{\varepsilon} \to \infty$ and $\chi(\varepsilon,\sigma) \ll D^{\varepsilon}\varepsilon_{xx}$. Correspondingly, from Eq. (1) follows

$$\varepsilon_t = D^{\varepsilon}\varepsilon_{xx}, \tag{4}$$

which corresponds to an ordinary diffusion equation (Fick's second equation), the solution of which for one dimension describes the propagation of a front, as is observed on the yield bench.

For parabolic hardening the situation is, evidently, reversed. In this case $D^{\varepsilon} \to 0$, and $\chi(\varepsilon,\sigma) \gg D^{\varepsilon}\varepsilon_{xx}$. From Eq. (1) we obtain

$$\varepsilon_t = \chi(\varepsilon,\sigma), \tag{5}$$

which can be interpreted as the condition of complete synchronization of the elementary acts of displacement in the entire volume of the sample. At the linear hardening stage it is impossible to analyze the behavior of autowaves on the basis of such a rough approach. Here it is necessary to take into account in detail the specific contribution of autocatalytic and damping factors to the kinetic process.

CONCLUSIONS

The experimental investigation of deformation localization in austenitic steel with a high nitrogen content made it possible to establish, as an elaboration of Ref. 27, that all autowave structures of plastic deformation can be observed in this material: a solitary front, a phase autowave, and a stationary dissipative structure. It was confirmed that a one-to-one relation exists between the law of work hardening and the character of the deformation localization (type of autowave process). It was found that the microscopic features, controlling the deformation process at the dislocation level (orientation of the slip planes, tendency toward transverse slip), is reflected in the character of the distribution of the localization bands of plastic deformation (macrolevel). It is obvious that this interrelation is realized directly via phenomena occurring at the mesoscopic level of plastic flow[28] and corresponding to the development and motion of the focus itself of the plastic deformation. Plastic deformation develops in a manner so that starting at the yield point and up to fracture, microscopic-scale elementary displacements self-organize into mesoscopic foci of deformation, and the latter in turn form a macroscopic pattern of deformation localization in the entire sample.

[1] I. A. Oding, V. S. Ivanova, V. V. Burdukskiĭ, and V. N. Geminov, *The Theory of Creep and Creep Strength of Metals* [in Russian], Metallurgizdat, Moscow, 1959, 488 pp.
[2] F. Garofalo, *Fundamentals of Creep and Creep-Rupture in Metals* [MacMillan, London, 1965; Metallurgiya, Moscow, 1968, 304 pp.].
[3] P. J. Wray, J. Appl. Phys. **41**, 3347 (1970).
[4] A. A. Presnyakov, *Localization of Plastic Deformation* [in Russian], Nauka, Alma-Ata, 1981, 119 pp.
[5] V. E. Panin, L. B. Zuev, V. I. Danilov, and N. M. Mnikh, Dokl. Akad. Nauk SSSR **308**, 1375 (1989) [Sov. Phys. Dokl. **34**, 940 (1989)].
[6] L. B. Zuev, V. I. Danilov, and N. V. Kartashova, JETP Lett. **60**, 553 (1994).
[7] L. B. Zuev, V. I. Danilov, and V. V. Gorbatenko, Zh. Tekh. Fiz. **65**(5), 91 (1995) [Tech. Phys. **40**, 456 (1995)].
[8] L. B. Zuev, N. V. Kartashova, V. I. Danilov *et al.*, Zh. Tekh. Fiz. **66**(11), 190 (1996) [Tech. Phys. **41**, 1189 (1996)].
[9] L. B. Zuev, Metallofizika (Kiev) **16**(10), 31 (1994).
[10] Yu. I. Chumlyakov, I. V. Kireeva, A. D. Korotaev, and L. S. Aparova, Fiz. Met. Metalloved. No. 2, 150 (1993).
[11] A. P. Bashchenko, A. V. Omel'chenko, and V. I. Soshnikov, Izv. Akad. Nauk. SSSR, Met. No. 4, 143 (1988).
[12] R. Jones and C. M. Wykes, *Holographic and Speckle Interferometry* [Cambridge University Press, Cambridge, 1983; Mir, Moscow, 1987, 328 pp.].
[13] A. Kadić and D. G. B. Edelen, *A Gauge Theory of Dislocations and Disclinations* [Springer-Verlag, Berlin, 1983; Mir, Moscow, 1986, 168 pp.].
[14] H. J. Kestenbach, Philos. Mag. **36**, 1509 (1977).
[15] L. B. Sujew, W. I. Danilow, S. A. Barannikova, and I. M. Kireewa, Z. Metallkd. **88**, 748 (1997).

[16] L. B. Zuev, V. I. Danilov, N. V. Kartashova, and S. A. Barannikova, Mater. Sci. Eng., A **324–326**, 699 (1997).
[17] L. B. Zuev and V. I. Danilov, Fiz. Tverd. Tela (St. Petersburg) **39**, 1399 (1997) [Phys. Solid State **39**, 1241 (1997)].
[18] H. Haken, *Information and Self-Organization* [Springer-Verlag, New York, 1988; Mir, Moscow, 1991, 240 pp.].
[19] E. C. Aifantis, Int. J. Plast. **3**, 211 (1987).
[20] Sh. Kh. Khannanov, Fiz. Met. Metalloved. , No. 4, 14 (1992).
[21] S. N. Nagornykh and G. F. Sarafanov, Metally, No. 3, 199 (1993).
[22] G. A. Malygin, Fiz. Tverd. Tela (St. Petersburg) **37**, 3 (1995) [Phys. Solid State **37**, 1 (1995)].
[23] V. A. Vasil'ev, Yu. M. Romanovskiĭ, and V. G. Yakhno, *Autowave Processes* (Nauka, Moscow, 1987).
[24] I. Prigogine and G. Nicolis, *Exploring Complexity* [W. H. Freeman, San Francisco, 1989; Mir, Moscow, 1990, 336 pp.].
[25] L. B. Zuev and V. I. Danilov, Int. J. Solids Struct. **34**, 3795 (1997).
[26] L. B. Zuev, Metallofizika (Kiev) **18**(5), 55 (1996).
[27] V. I. Danilov, S. I. Barannikova, L. B. Zuev, and I. V. Kireeva, Fiz. Met. Metalloved. **83**, 140 (1997).
[28] V. E. Panin, Uzv. Vyssh. Uchebn. Zaved. Fiz., No. 1, pp. 7–34 (1998).

Translated by M. E. Alferieff

Silicon two-emitter differential tensotransistor with an accelerating electric field in the base

G. G. Babichev, S. I. Kozlovskiĭ, V. A. Romanov, and N. N. Sharan

Institute of Semiconductor Physics, Academy of Sciences of Ukraine, 252650 Kiev, Ukraine
(Submitted May 6, 1998; resubmitted March 2, 1999)
Zh. Tekh. Fiz. **69**, 63–68 (October 1999)

Results are presented from a study of silicon two-emitter tensotransistors. Tensotransistors are strain-sensitive semiconductor devices with a horizontal structure and an internal differential output. The optimum device topology and its main characteristics are determined. The transfer coefficient of the transistor is close to unity. © *1999 American Institute of Physics.* [S1063-7842(99)00910-1]

Integrated semiconductor pressure sensors occupy a secure leading position in the manufacture and utilization of mechanical sensors. The existing semiconductor sensors with a resistive output can be divided into two groups: ''monopolar'' and ''bipolar,'' in which the output signal is formed by the directed motion of majority and minority charge carriers, respectively. The first group includes piezoresistive sensors and tenso-emf sensors (utilizing the mechanical analog of the Hall effect), and the second group consists of sensors with sensitive elements based on bipolar semiconductor devices: multicollector tensotransistors and tensothyristors.[1–3] One of the most important characteristics of a pressure sensor is the relative current sensitivity,

$$S_R=\left[I_0^{-1}\left(\frac{\partial(I)}{\partial\sigma}\right)_{\sigma=0}\right]. \tag{1}$$

Here $I(\sigma)=I_0+\Delta I(\sigma)$, where I_0 and $\Delta I(\sigma)$ are, respectively, the current passing through the sensor in the absence of pressure and the change in the current owing to a pressure and σ is the mechanical stress in the region where the sensor lies. S_R is relatively small for monopolar sensorss, since the relative change in the current during deformation, which, in turn, is proportional to the relative change in the mobility of the majority charge carriers, is only a few percent. For bipolar semiconductor sensors, even a small change in mobility can lead to a large change in the current, $\Delta I(\sigma)$. This is because $\Delta I(\sigma)$ in this case is proportional both to the relative change in the mobility and to the ratio of the magnitudes of the pulling electric field E and the diffusion electric field E_D.[4] The product of these parameters can be quite large, even for a small change in mobility because of the factor E/E_D. Because of this, the relative sensitivity of bipolar sensors is substantially higher than that of monopolar sensors. Thus, for two-collector tensotransistors, S_R, which is given by[1]

$$S_R=\left[(I_c)^{-1}\left(\frac{\partial(I_{c1}-I_{c2})}{\partial\sigma}\right)_{\sigma=0}\right] \tag{2}$$

($I_c=I_{c1}+I_{c2}$, where I_{c1} and I_{c2} are the individual collector currents), is more than an order of magnitude higher than the analogous value for monopolar sensors. However, S_R, which is also known as the conversion efficiency,[5,6] neglects the current consumption of the entire tensotransistor as a whole. Thus, by analogy with magnetotransistors,[5] we can introduce the concept of the conversion efficiency of tensotransistor, S_R^T, defining it as

$$S_R^T=\left[I_{\text{tot}}^{-1}\left(\frac{\partial(I_{c1}-I_{c2})}{\partial\sigma}\right)_{\sigma=0}\right], \tag{3}$$

where I_{tot} is the sum of all the currents required to bias the tensotransistor.

Equation (3) can also be regarded as a special case of Eq. (1) for a two-collector tensotransistor.

In the tensotransistors studied before,[1–3] in order to achieve maximum sensitivity (through the factor E/E_0) we included a separating $p-n$ junction that separates the base region of the tensotransistor from the substrate. Since the thickness of the base of the tensotransistor was much less than the diffusion mixing length for the current carriers, the bulk of the injected carriers did not reach the collector $p-n$ junctions, but were collected by the separating $p-n$ junction (the collector of a parasitic vertical transistor) and recombined near the current electrode of the base that was most distant from the emitter. For these tensotransistors, I_{tot} equals the sum of the currents flowing through the base electrodes I_{bb}, the emitter I_e, and the separating $p-n$ junction of the tensotransistor, I_s:

$$I_{\text{tot}}=I_{bb}+I_e+I_c+I_s. \tag{4}$$

With Eqs. (3) and (4), Eq. (2) can be written in the form

$$S_R^T=\frac{(\alpha_{c1}+\alpha_{c2})S_R}{1+(\alpha_{c1}+\alpha_{c2})(1+I_{bb}/I_c+I_s/I_c)}, \tag{5}$$

where $\alpha_{c1(c2)}$ are the transfer coefficients for the $c1(c2)$ collector currents, which are related to the efficiencies of the emitter γ_e and collectors $\gamma_{c1(c2)}$, as well as to the transfer coefficient $\beta_{c1(c2)}$, by

$$\alpha_{c1(c2)}=\gamma_e\beta_{c1(c2)}\gamma_{c1(c2)}, \tag{6}$$

where $\beta_{c1(c2)}$ is the ratio of the hole components of the emitter I_{pe} and collector $I_{c1(c2)}$ currents, i.e.,

$$\beta_{c1(c2)}=I_{pe}/I_{c1(c2)}. \tag{7}$$

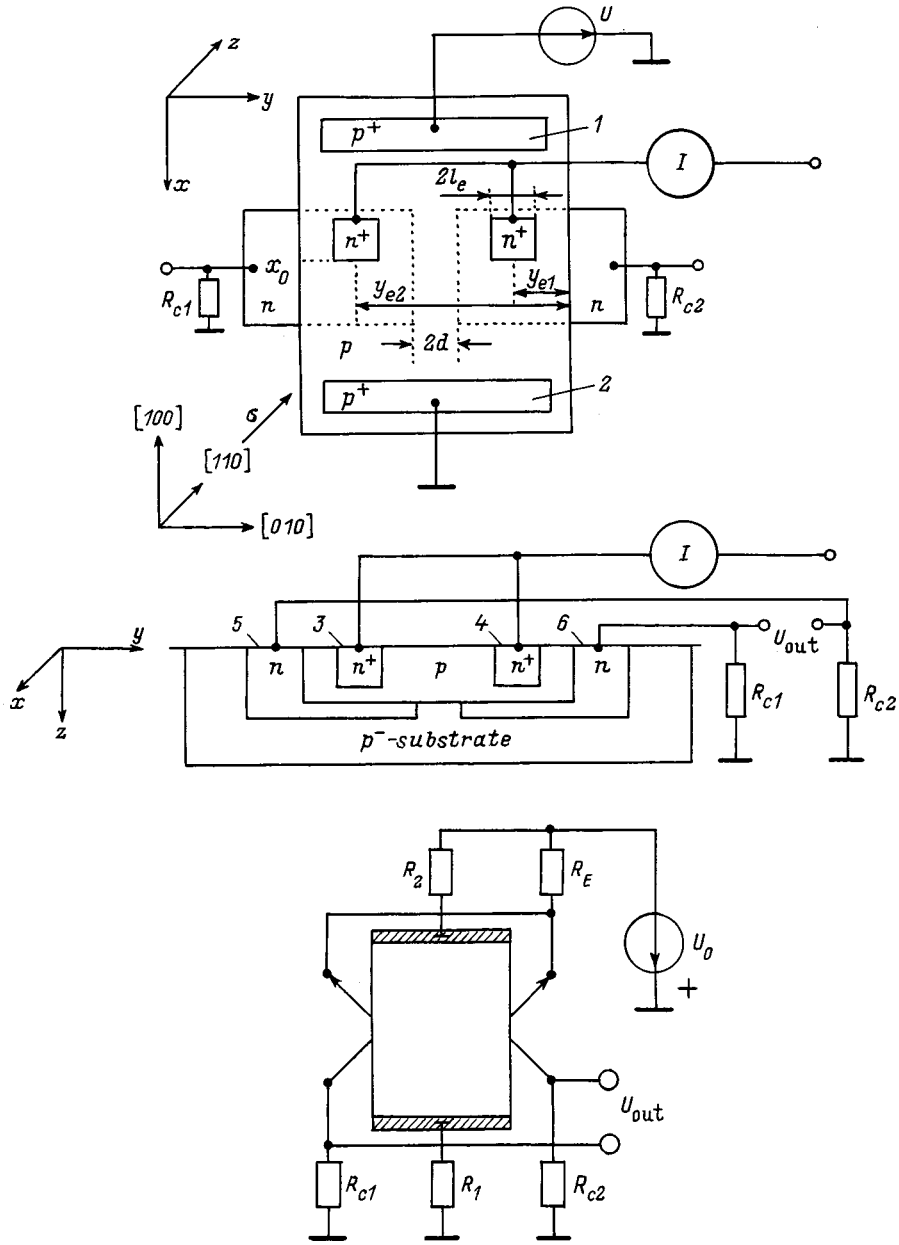

FIG. 1. Topology of the tensotransistors: *1, 2* — base electrodes; *3,4* — emitters; *5,6* — collectors C_1 and C_2, respectively; σ is the direction of uniaxial compression.

When a separating $p-n$ junction is present and recombination takes place in the base region of a tensotransistor, $I_{pe} \gg I_{c1(c2)}$, so that for the tensotransistors considered previously, $\beta_{c1(c2)}$ and the associated $\alpha_{c1(c2)}$ are considerably smaller than unity. (See Refs. 1 and 7, for example.)

In the optimum case ($I_{bb}, I_s \ll I_c$, $\alpha_{c1(c2)} \geqslant 1$), the conversion efficiency and the relative current sensitivity of a tensotransistor are essentially the same. However, in practice, $\alpha_{c1(c2)} \ll 1$ and $S_R \neq S_R^T$. In this case, it is appropriate to optimize the design, topology, and operating regime of the tensotransistors simultaneously with respect to the two parameters S_R and S_R^T, thereby reaching both a maximum relative sensitivity and a minimum energy consumption for the device.

In this paper we analyze the operation of a two-emitter differential tensotransistor with higher S_R and S_R^T owing to its design features, which eliminate the need to create a separating $p-n$ junction ($I_s = 0$) and make it possible to achieve substantially higher transfer coefficients for the collector currents.

The topology and electrical circuit diagram for connecting a tensotransistor are shown in Fig. 1.

Nonequilibrium charge carriers are injected into the base region of a tensotransistor by two emitters whose *n*-type regions are at different potentials. The current carriers drift along the base region of the tensotransistor in the accelerating electric field at the base, E_x. (E_x is determined by the difference in the potentials U at the p^+-base electrodes.) In the absence of elastic deformation, the collector currents owing to diffusive flow of charge carriers in the y and z directions, are equal. Because of this, the output signal U at the load resistances R_{c1} and R_{c2} is zero.

Under uniaxial elastic deformation of the base region of tensotransistors in the [110] crystallographic direction, the majority carrier (hole) mobility becomes anisotropic in the xy plane, so a transverse potential difference develops which

is shifted in the forward direction or in the reverse direction by one of the emitter junctions, depending on the sign of the deformation. The resulting modulation in the injection of emitter $p-n$ junctions leads to an imbalance in the currents in the collector circuits of the tensotransistor and to the appearance of an output signal.

The operating principle of the tensotransistors described here is formally equivalent to the principle of the differential amplifier magnetic sensor,[5] in which the injection of emitter $p-n$ junctions is modulated by the Hall field.

As before,[1–3] we shall assume that the tensotransistor is located on a membrane which transforms a load uniformly distributed over its surface into a uniaxial elastic tension (or compression) deformation of its base region in the [110] crystallographic direction. In this case, the anisotropy parameter a is related to the elastic mechanical stress σ in the base of the tensotransistor by

$$a=\Pi_{44}\sigma/2, \tag{8}$$

where Π_{44} is the shear piezoresistive coefficient for p-Si.

In order to optimize the topology and geometric dimensions of the elements of a tensotransistor, we shall calculate its main characteristics. The calculation will be done in the approximations of a low level of injection and weak deformations, such that the anisotropy parameter is rather small, i.e., $|a|\ll 1$. In view of the smallness of $2d/l_y\ll 1$, we shall also neglect the effect on the spatial distribution of the nonmajority charge carriers in the base of the tensotransistor owing to the layer of thickness $2d$ (Fig. 1) that separates the collector $p-n$ junctions.

We limit the base of the tensotransistor to $0<x<l_x$, $0<y<l_y$, and $0<z<l_z$. Under the above assumptions, the equations for the distribution of the electrical potential in the base region of the tensotransistor are the same as before.[1] Thus, the results obtained there on the choice of geometric dimensions for the base of the tensotransistor and the position of the emitters remain valid: $l_x/l_y=0.2$ and $x_c/l_x\geqslant 0.25$, where x_c is the distance from the center of the emitters to the nearest base electrode.

For calculating the output characteristics of the tensotransistor, we must know the spatial distribution of the nonequilibrium charge carrier concentration in the base of the tensotransistor. Using the source function[2] $n(x,y,z,x',y',z')$, we find an expression for the concentration of the nonequilibrium carriers by integrating over the volume V_c emitters:

$$n(x,y,z)=\int dx'\int dy'\int dz'\rho(x',y')n(x,y,z,x',y',z'),$$

$$x',y',z'\in V_e. \tag{9}$$

Here $\rho(x',y')$ is the source density of the nonequilibrium carriers, which in our case is a function of the pulling field, the anisotropy parameter, and position

$$\rho(x',y')=n_0\{\exp[\Phi(x',y')]-1\}, \tag{10}$$

where

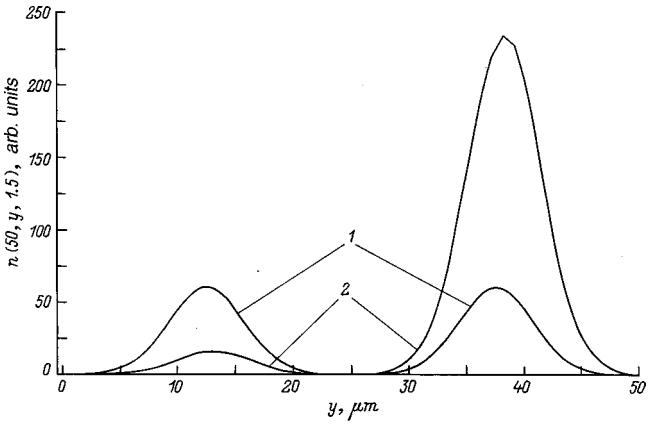

FIG. 2. Transverse distribution of the nonequilibrium charge carrier distribution for different values of the anisotropy parameter α: *1* — 0, *2* — 0.055.

$$\Phi(x',y')=e\frac{\varphi_0+E_x x'+aE_x\left(y'-\dfrac{l_y}{2}\right)}{k_BT}, \tag{11}$$

φ_0 is the initial bias voltage on the emitter $n-p$ junction, T is the temperature, k_B is the Boltzmann constant, and e and n_0 are the electronic charge and the equilibrium electron concentration in the base.

The following analysis is for dimensions (all in μm) of the base $l_x\times l_y\times l_z=200\times 50\times 3$ and of the emitters $l_x'\times l_y'\times l_z'=4\times 5\times 1.5$, a separation $y_{e1}-y_{e2}=25$ between the centers of the emitters, $x_0=42$, and a diffusion mean free path for the nonequilibrium charge carriers of $L=30$. The pulling field is $E_x=500$ V/cm, and the surface recombination rate $S=0$. In the calculations, we have taken an anisotropy parameter of $a=0.055$, which corresponds to a uniaxial elastic mechanical stress $\sigma=800$ kgf/cm^2. The changes in each specific case are indicated specially. Note that the condition of a low level of injection imposes limits on the quantity

$$y_{e1}-y_{e2}+l_y'\leqslant\frac{2k_BT}{|a|E_x}\ln\left[\frac{n_0}{p_0}+1\right]. \tag{12}$$

Here p_0 is the equilibrium concentration of the majority carriers (holes) in the base of the tensotransistor. In the opposite case, the nonequilibirum carriers generated by an $n-p$ junction will short out the transverse field and the modulation in the injection vanishes. Figure 2 shows transverse distributions of the concentration of nonequilibrium charge carriers in the base of a tensotransistor along the line specified by the coordinates, μm $(80,y,3)$. Evidently, in the absence of deformation ($\sigma=0$), the distribution is mirror symmetric with respect to the central plane $(x,l_y/2,z)$ (curve 1). When a mechanical stress σ is present, the distribution becomes sharply anisotropic with the concentration increasing rapidly at one of the collectors and decreasing at the other (curve 2). Here the overall number of nonequilibrium carriers increases, as implied by a comparison of the areas under curves 1 and 2. This variation in the spatial distribution of the nonequilibrium electrons in the base is a direct consequence of the modulation in the injection.

The collector currents are found by taking the integral

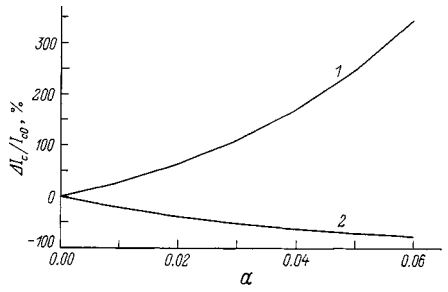

FIG. 3. Variation in the C_1 (*1*) and C_2 (*2*) collector currents as functions of the anisotropy parameter.

$$I_{c1(c2)} = -eD_n \int_{x_0}^{1} dx \left[\int_0^1 \left. \frac{\partial n}{\partial y} \right|_{y=0,(l_y)} dz + \int_{0,\left(\frac{l_y}{2}+d\right)}^{\frac{l_y}{2}-d,(l_y)} \left. \frac{\partial n}{\partial z} \right|_{z=l_z} dy \right]. \quad (13)$$

Figures 3 and 4 show the relative variations in the collector currents as functions of the anisotropy parameter and pulling electric field in the tensotransistor base. These figures imply that when elastic deformation forces are applied, the collector current can increase by more than a factor of 3, and its dependence on the anisotropy parameter (on the amount of deformation) is substantially nonlinear.

Figure 5 shows the variations in the components I_{cz} and I_{cy} of the collector currents flowing through the (x,y,l_z) and $(x,0,z)(x,l_y,z)$ planes of the collector $p-n$ junctions. Evidently, the main contribution to l_c is from the current flowing through the (x,y,l_z) plane of the collector junction. This curve also implies that for this tensotransistor, the relative increase in the collector current during elastic deformation of its base region is roughly a factor of 2 smaller than for the $n-p-n$ tensotransistor studied before under the same conditions,[1] since in this design, the charge carriers injected by the emitter, whose current increases during deformation, reach both collector junctions. This effect is substantially less for an $n-p-n$ tensotransistor, since its collectors are separated from one another by a substantially greater distance.

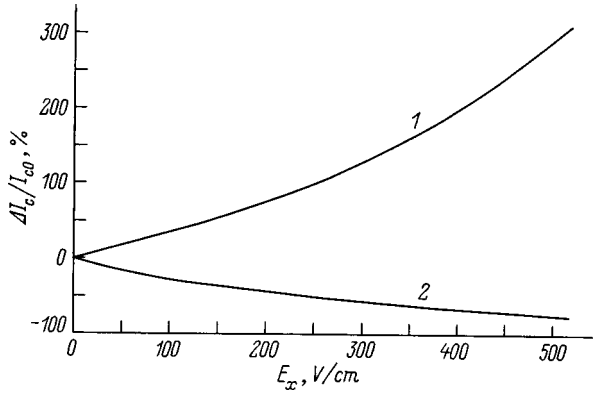

FIG. 4. Field dependence of the relative variation in the C_1 (*1*) and C_2 (*2*) collector currents.

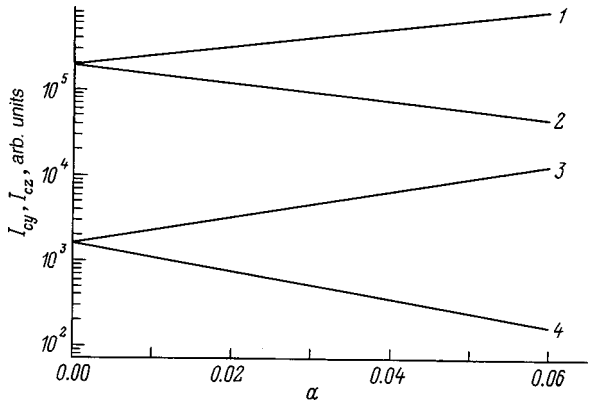

FIG. 5. Dependence of the *z* (*1,2*) and *y* components (*3,4*) of the C_1 (*1,3*) and C_2 (*2,4*) collector currents on the anisotropy parameter.

The distribution of the collector ion currents $j_{c1}(x)$ and $j_{c2}(x)$ along the lines specified by the coordinates (x,y_{e1},l_z) and (x,y_{e2},l_z), respectively, are shown in Fig. 6. The length of the collector junctions can be optimized on the basis of this dependence. In our case, the collector junctions can be bound by $40<x<55$.

The relative variation in the collector currents during deformation are plotted as a function of the distance between the collectors in Fig. 7. As the distance between the collectors increases, there is a rise in the transverse potential difference that modulates the local injection at the emitter $p-n$ junctions.

The results of a study of the relative variation in the collector currents, $\Delta I_c/I_{c0}$ ($\Delta I_c = I_c - I_{c0}$, where I_c and I_{c0} are the collector currents with and without deformation, respectively), as a function of the diffusion displacement length L for the charge carriers and the width of the emitter $p-n$ junctions reduce to the following: when L is reduced from 30 to 5 μm, $\Delta I_c/I_{c0}$ increases slightly, while I_{c0} decreases by a factor of 4.5, which leads to an undesirable increase in the output resistance of the device. Similar behavior is typical in the dependence of $\Delta I_c/I_{c0}$ on the width of the emitter $p-n$ junctions when the distance between

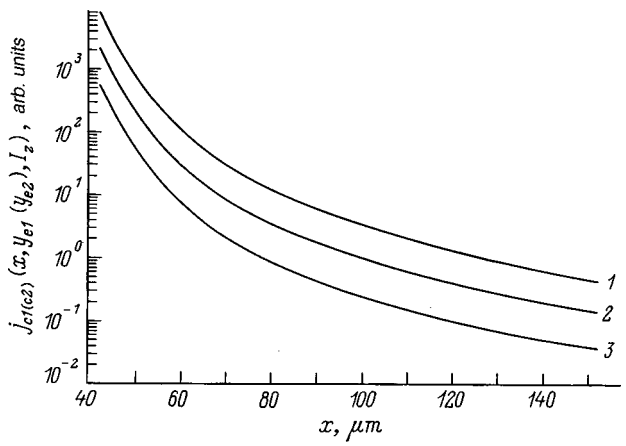

FIG. 6. Longitudinal distributions of the collector current densities without (*2*) and with (*1,3*) deformation of collectors C_1 (*2*) and C_2 (*1,3*).

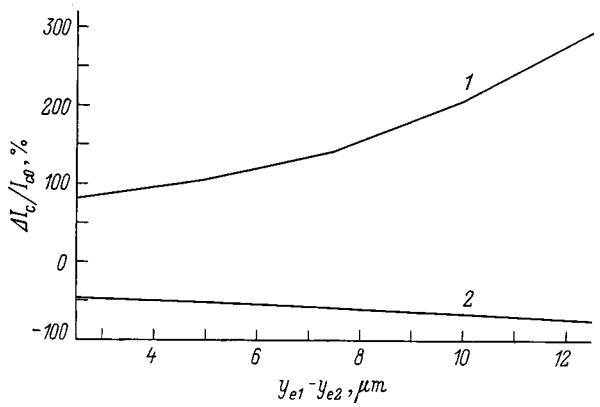

FIG. 7. Relative variations in the C_1 (*1*) and C_2 (*2*) collector currents with deformation as functions of the distance between the collectors.

their centers, $y_{e1}-y_{e2}$, is held constant. As the width of the emitter junctions is reduced, $\Delta I_c/I_{c0}$ increases slightly, while I_{c0} increases substantially.

We conclude by estimating the magnitudes of the relative S_R and absolute S_A sensitivities, as well as the transfer coefficient of the tensotransistor. The curves of Fig. 3 imply that $S_R=1.8\times10^{-3}\,\mathrm{cm^2/kgf}$. An expression for calculating the absolute sensitivity can be obtained from Eq. (2) by replacing $(I_c)^{-1}$ by the load resistance in the collector circuits, $R_c=R_{c1}=R_{c2}$. Then, for zero pressure collector currents $I_{c1}=I_{c2}=1\,\mathrm{mA}$ and $R_c=5\,\mathrm{k\Omega}$, we obtain $S_A=18\,\mathrm{mV\cdot cm^2/kgf}$. For a pulling electric field in the base of $E_x=500\,\mathrm{V/cm}$, Eq. (7) yields $\beta_{c1(c2)}=0.8$, which is roughly two orders of magnitude higher than for the earlier tensotransistor designs.

We thank Z. S. Gribnikov for stimulating these studies.

[1] G. G. Babichev, V. N. Guz', I. P. Zhad'ko *et al.*, Fiz. Tekh. Poluprovodn. **26**, 1244 (1992) [Sov. Phys. Semicond. **26**, 694 (1992)].
[2] S. I. Kozlovskiĭ, Fiz. Tekh. Poluprovodn. **29**, 1244 (1995) [Semiconductors **29**, 930 (1995)].
[3] G. G. Babichev, V. N. Guz', I. P. Zhad'ko *et al.*, Fiz. Tekh. Poluprovodn. **26**, 1723 (1992) [Sov. Phys. Semicond. **26**, 964 (1992)].
[4] I. I. Boĭko and V. A. Romanov, Fiz. Tekh. Poluprovodn. **11**, 817 (1977) [Sov. Phys. Semicond. **11**, 481 (1977)].
[5] H. P. Baltes and R. S. Popovic, Proc. IEEE **74**, 1107 (1986).
[6] S. Takamiya and K. Fujikawa, IEEE Trans. Electron Devices **ED-19**, 1085 (1972).
[7] S. I. Kozlovskiĭ, Fiz. Tekh. Poluprovodn. **30**, 1544 (1996) [Semiconductors **30**, 810 (1996)].

Translated by D. H. McNeill

Effect of prolonged operation and temperature on the spectra of silicon carbide light-emitting diodes operating in an electric breakdown regime

A. M. Genkin, V. K. Genkina, and L. P. Germash

Kiev Polytechnical Institute Ukrainian National Technical University, 252056 Kiev, Ukraine
(Submitted April 13, 1998)
Zh. Tekh. Fiz. **69**, 69–76 (October 1999)

The effect of prolonged operation and temperature on the spectral distribution of the quantum yield of silicon carbide light-emitting diodes operating in an electric breakdown regime is investigated. The spectrum of the light-emitting diodes is found to change in a regular manner during prolonged operation. It is found that self-absorption and the temperature shift of the bands have a large effect on the spectral distribution of the temperature coefficient of emission. The conditions for obtaining the most stable light-emitting diodes possible, which can be used as standard visible-range emitters, are determined. © *1999 American Institute of Physics.* [S1063-7842(99)01010-7]

Silicon carbide light-emitting diodes operating in an electric breakdown regime (breakdown LEDs) meet, to the maximum degree possible, the requirements for standard sources of optical radiation for sensitivity calibration of spectral apparatus. They are distinguished by a wide flat spectrum similar to the spectrum of a black body at 6000–8000 K, high temperature and temporal stability of the integrated radiation power, and resistance to current overloads and corrosive actions.[1]

One of the basic parameters whose values determine the possibility of using breakdown LEDs as standard sources are the temperature and temporal instability of the spectral distribution of the quantum yield of radiation within the operating wavelength range (250–800 nm). However, previous investigations of the instability of the quantum yield of breakdown LEDs[2,3] concern primarily its integrated values (over the wide spectral range). In addition, the radiation was extracted through a base silicon carbide crystal, since self-absorption strongly influences the characteristics of the devices.

In the present work we investigated the effect of prolonged operation and temperature on the spectral distribution of the quantum yield of silicon carbide LEDs operating in the electric-breakdown regime with radiation extracted through a thin p region.

The p–n structures were prepared by alloying Silumin disks with commercial SiC–6H crystals grown by the Leli method. The structures were 150–300 μm in diameter. Fusing-in was done in the faces of the crystals perpendicular to the crystallographic axis **C**. The uncompensated donor density in the initial crystals was of the order of 10^{18} cm^{-3}.

Breakdown in the experimental structures was localized in microplasmas, whose position and characteristics depended on the degree of doping, the surface treatment method used for the initial crystals, the temperature–time regime of alloying, the masses of the Silumin disks, and other factors. Structures in which the microplasmas uniformly fill the entire area or are localized within a ring occupying a substantial portion of the p–n junction area were selected. Such samples were obtained by careful optimization of the technological regimes with defects artificially introduced into a near-surface layer of the initial silicon carbide wafers. In the optimal variant the microplasmas are switched on at essentially the same voltages. In the current–voltage characteristic, an inflection preceding the section of sharp current growth corresponded to the appearance of microplasmas. The high stability of the radiation power at constant current during prolonged operation was achieved for such samples on a section of ''sharp'' breakdown located at currents quite far above the inflection point.[2] Some characteristics of the breakdown LEDs containing similar structures are presented in Ref. 1.

The maximum number of microplasmas was switched on when the reverse current pulse through the experimental structures reaching several amperes. These microplasmas became equally bright and a system could be discerned in their arrangement. They were localized predominantly along closed segments of curves surrounding the region where the maximum brightness of injection electroluminescence, arising when a sample is powered with dc pulses with amplitude greater than 0.5–1 A was observed. For this reason, it can be inferred that the appearance of microplasmas is due to regular nonuniformities of the doping front. The number density of the microplasmas increased with the degree of doping of the initial crystals.

Some of the experimental p–n structures formed groups of structures together, approximately 0.5 mm apart, on the same crystal. Such structures exhibited characteristics with the minimum variance, which makes it possible to attribute the uncontrollable technological factors affecting them appreciably to the properties of the initial SiC–6H wafers and the temperature–time regime of alloying.

In order to use the photons whose energy is greater than the band gap in SiC, the radiation of the LEDs was extracted through a thin p region. Such photons are completely absorbed when the radiation is extracted through the bulk of

the SiC. As shown in the present work, the temperature dependence of the optical absorption in this case can strongly influence device stability.

The presence of drops of the emitter alloy made it possible to obtain, by thermal compression welding of an aluminum wire, a reliable stable contact to the *p* region, capable of passing for more than 1000 h a pulsed current with amplitude exceeding 10 A without appreciable degradation. The resistance of such a contact did not exceed several Ω. When radiation was extracted from the *p*-region side the contact ''masked'' no more than 20–30% of the emitting area of the *p–n* junction with a uniform field of microplasmas and, for all practical purposes, did not impede the extraction of radiation from the samples with ring-shaped radiation topography. The simple possibility of obtaining a ''transparent'' stable contact to the *p* region is a substantial advantage of the alloying technology for *p–n* junction fabrication.

The structures on which the measurements were performed were placed in heat-removing cases. The base contact was obtained by vacuum deposition of Ti + Ni.

The method for determining the temporal instability of the quantum yield η consisted of measuring the signal I from a photomultiplier when the photomultiplier is irradiated with radiation from the experimental LED before and after the given operating periods with a constant relative arrangement of the periods and comparing this signal with the signal I_e from a reference emitter. In so doing, the relative radiation intensity $\Phi = I/I_e$ of the experimental LEDs which is proportional to the quantum yield, was found. The relative change in the quantum yield over a period t of operation of the LED was determined as $\delta\eta = [\Phi(0) - \Phi(t)]/\Phi(0)$.

A set of breakdown LEDs of the same type as the experimental LEDs, which had undergone preliminary aging and were subjected to mutual verification throughout the entire period of the tests, was used as reference emitters.[4] The reference LEDs were placed together with the experimental LEDs on a common removable holder in front of the photomultiplier unit about 30 cm from the photocathode in a massive case. The arrangement of the LEDs in the holder was not changed during the tests. Its construction made it possible to operate simultaneously all of the experimental LEDs and to control their characteristics. For this, the holder was moved from one measuring stand to another. Radiation in a narrow spectral interval was extracted using interference light filters, placed in front of the photocathode of the photomultiplier. Reference LEDs were used under the same conditions as the measured LEDs. Their radiation intensity was the same. The same powering and switching systems were used to switch them on.

Measurements in the spectral range 687–361 nm were performed in the photon-counting mode using an FÉU-84 photomultiplier. An FÉU-142 photomultiplier, which is insensitive to visible-range radiation, was used in the UV region. The photomultipliers were carefully selected from a large number of models from different batches according to the criteria of maximally stable detection of weak radiation fluxes (intensity of the order of 10^5–10^7 photons per second) in the single-electron regime. It should be noted that the FÉU-142 photomultiplier had a multiplication system with a quite high gain for efficient operation in the single-electron mode. However, a model which was substantially different from the others and had excellent characteristics was found. The FÉU-142 signal was recorded in integrated light (no light filters).

As a rule, in the temperature range 300–400 K the curves of the spectral density of the quantum yield versus temperature were linear. This makes it possible to characterize them by the relative temperature coefficient $\gamma_{h\nu}$, expressed in %/K.

To obtain the spectral dependence $\gamma_{h\nu}(h\nu)$ the spectrum of a sample was measured successively at room temperature T_1 and at a temperature T_2 100 K above room temperature. In so doing, it was checked that the photosensitivity of the detecting system remained constant. The quantity $\gamma_{h\nu}(h\nu)$ was determined for each photon energy with step $\Delta(h\nu) = 0.05$ eV according to the formula

$$\gamma_{h\nu}(h\nu) = 100\%[I(T_2) - I(T_1)]/I(T_1)(T_2 - T_1)].$$

The spectrum $N_{h\nu}(h\nu)$ of a sample, expressed in units of the number of emitted photons in a unit photon energy range, and the spectral distribution of the relative slope of the spectrum $\xi(h\nu) = (1/N_{h\nu}) \cdot \Delta N_{h\nu}/\Delta(h\nu)$ were used for combined analysis.

The radiation spectra were measured in the photon energy range 1.8–3.8 eV with a 0.05 eV step using a ZMR-3 monochromator. A photon counter based on an FEU-79 photomultiplier, carefully selected to ensure the maximally stable radiation detection, was used as the photodetector. The experimental samples were placed in a special holder directly in front of the input slit of a monochromator. The holder made it possible to regulate and stabilize the sample temperature in a range from room temperature up to 400 K. The magnitudes of the signal from the experimental sample within the entire spectral range during the measurements of two spectra corresponding to room and a high temperature of the sample (about 2 h), were reproduced with an error not exceeding 0.1%.

Fourteen typical *p–n* junctions, some of which (containing the same first number in their designation) were placed on the same crystal, were selected for measurements of the temporal instability. Multiplexed LEDs with several individually powered emitting *p–n* junctions in the same case were prepared from such structures. The LEDs were placed on two holders. Each holder contained two standard emitters, one of which was used as the main emitter (the signals of all other samples were normalized with respect to it) and the other served as a control.

Light-emitting diodes were operated ''around the clock'' at a 50 mA current. The temperature of the holder was about 350 K. This ''easy'' (as compared with Ref. 3) powering regime is optimal for device operation on considerations of moderate power dissipation.

The relative radiation intensity was measured in narrow spectral intervals with a current of 50 mA through the LED as well as in integrated light at currents of 3, 10, and 50 mA using an FÉU-84 photomultiplier (whose spectral sensitivity range is 300–740 nm) and 50 mA with an FÉU-142 photomultiplier (whose spectral sensitivity range is 112–365 nm

TABLE I.

	U, V			$(\delta\eta)_{max}$, %									
				Integrated light				Interference light filters					
				FÉU-84			FÉU-142						
	Current, mA							Wavelength, nm					
Sample	3	10	50	3	10	50	50	361	410	474	560	618	687
Holder 1													
31–1	20.6	21.2	22.8	7	10	10	15	8	6	7	11	12	11
31–2	20.4	21.0	22.4	10	10	8	21	7	6	6	9	11	9
21–1	18.5	19.0	20.3	4	6	6	9	8	7	2	5	6	4
21–2	18.5	18.9	20.0	2	5	6	8	9	6	5	4	5	3
19–1	18.4	18.8	19.5	1	5	7	9	12	11	9	5	5	3
19–2	18.2	18.6	19.7	−10	2	5	8		9	7	3	3	1
19–3	18.4	18.9	19.9	2	7	9	13	15	12	10	6	6	4
19–4	18.1	18.6	20.0	−14	0	0	6	8	7	6	3	3	1
21–3	18.6	19.0	19.8	−1	7	10	20		13	12	8	9	4
26–3	17.3	20.5	21.6	−2	1	2	2	1	3	4	2	2	0
Holder 2													
11	18.0	18.5	19.4	−4	−5	9	37	16	9	5	3	2	1
14	20.4	20.7	21.5	6	4	4	17	12	10	8	5	4	3
28–1	22.5	23.4	25.7	8	10	9	25	10	6	5	6	9	5
28–2	22.3	23.1	25.2	8	7	6	17	4	2	3	5	7	2
18–5	21.4	22.4	23.3	15	16	14	19	18	14	12	11	12	12
18–3	21.1	21.6	22.8	1	0	−1	0	−1	−1	−1	−1	−1	−1

with a relative spectral sensitivity of the photocathode at 365 nm corresponding to 0.4% according to the manufacturer's certificate for this sample).

For all samples the voltage corresponding to the inflection in the current–voltage characteristic preceding the section of sharp current growth ranged from 17 to 21 V. The current corresponding to the inflection did not exceed 1 mA.

Table I gives data characterizing the form of the current–voltage characteristic of the experimental samples in the 3–50 mA section as well as the maximum relative change $(\delta\eta)_{max}$ in quantum yield over the entire operating period (5779 h for samples on holder 1 and 5965 h on holder 2).

As one can see from Table I, the 3–50 mA section corresponded to ''sharp'' breakdown. Considering its distance from the inflection in the current–voltage characteristic, one would expect, as follows from the preceding investigations,[3] a high temporal stability of the integrated quantum yield of light-emitting diodes on this section.

The samples 26–3 and 18–3 were not subjected to continuous operation; they were used as control samples. The measurement error can be judged according to the value of $(\delta\eta)_{max}$ corresponding to them. As one can see from Table I, $(\delta\eta)_{max}$ for 18–3 did not exceed 1%, which gives an estimated error of 1% in determining $\delta\eta$ for samples placed on holder 2. The values of $(\delta\eta)_{max}$ for the control sample 26–3 turned out to be higher [$(\delta\eta)_{max}$ reaches 4% in one case], attesting to an unsuccessful choice of the standard LEDs for holder 1. Apparently, one or both of them exhibited instability during the testing period, even though they were not subjected to continuous operation. For this reason, we estimate the error in determining $\delta\eta$ for holder 1 to be 4%. It is interesting to note that the random character of the quantity $(\delta\eta)_{max}$, corresponding to the control sample 26–2, obtained for various spectral intervals (apparently, temporal ''pulsations'' of the quantum yield appeared). A large change was not observed in the current–voltage characteristic of the main and control reference LEDs during the period of the tests. Unfortunately, this does not make any of them preferable with respect to stability.

As one can see from Table I, the quantum yield, as a rule, increases as a result of continuous operation. The maximum value $(\delta\eta)_{max}$, reaching 37% for one of the samples, is observed in the UV region of the spectrum (integrated light with the FÉU-142). Substantially lower values of $(\delta\eta)_{max}$ were obtained in the spectral range 361–687 nm; they do not prevent using the devices as standard emitters. The integrated quantum yield corresponding to currents of 3 and 10 mA was also quite stable. The higher negative values of $(\delta\eta)_{max}$ with 3 mA for samples 19–2 and 19–4 agree with the character of the change in their current–voltage characteristic, similar to that described previously in Ref. 3. These samples are distinguished by lower values of the operating voltage and, correspondingly, a position closer to the inflection in the current–voltage characteristic.

The spectral dependence of $(\delta\eta)_{max}$ averaged over all samples is displayed in Fig. 1. The circles denote the averages, and the short horizontal bars denote the maximum and minimum values of $(\delta\eta)_{max}$. The data on $(\delta\eta)_{max}$ in the UV range (integrated light with the FÉU-142) are also plotted there. Since (when the form of the emission spectrum of LEDs and the photosensitivity spectrum of the FÉU-142 are taken into account) photons with wavelengths from a small region near 300 nm make the main contribution to the signal,

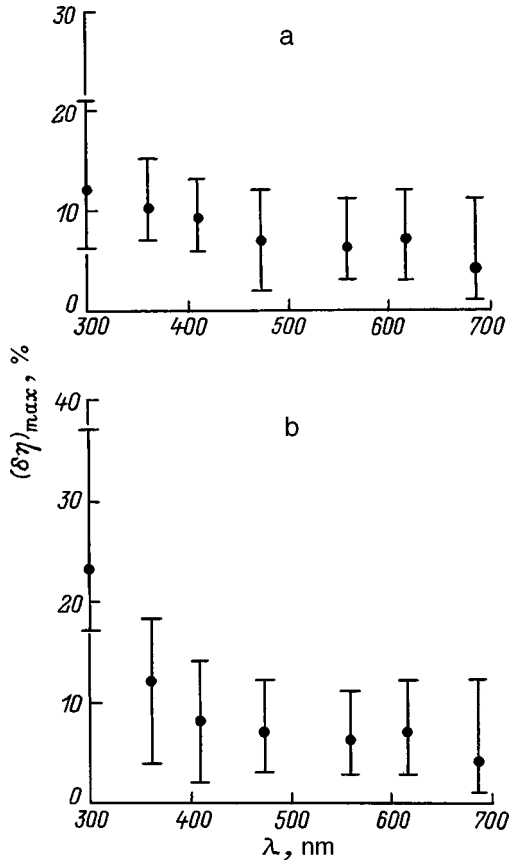

FIG. 1. Spectral dependence of the maximum relative change in quantum yield of breakdown LEDs over the entire operating period: a — holder 1, b — holder 2.

a wavelength of 300 nm was conditionally assigned to the corresponding value of $(\delta\eta)_{\max}$.

As one can see from the figure, the change in the spectrum of the LEDs as a result of their continuous operation is of a regular character. It should be noted that for spectral curves of the samples located on holders 1 and 2, not only are the characters of the curves close, but the values of $(\delta\eta)_{\max}$ are close. A large difference is observed only at 300 nm. It seems to be of a random character and is due to the stronger influence of uncontrollable technological factors on the characteristics of samples in the UV region of the spectrum as compared with the visible region. A similar change in the shape of the emission spectrum of breakdown LEDs (spectra before and after continuous operation were compared, in both cases normalized to 1 for the same wavelength), undergoing continuous operation with the maximum possible current load with radiation extraction through the base silicon carbide crystal was noted in Ref. 3.

The dependence of Φ averaged over all samples on the operating time t is presented in Fig. 2. As one can see, the quantum yield increases monotonically during the operating time in most cases. For holder 1 a less sharp dependence is observed, possibly because of the higher instability of the standard LEDs.

In the course of the tests, the values of the voltage on the sample were recorded in parallel with the measurement of the radiation intensity. It was established that for all samples

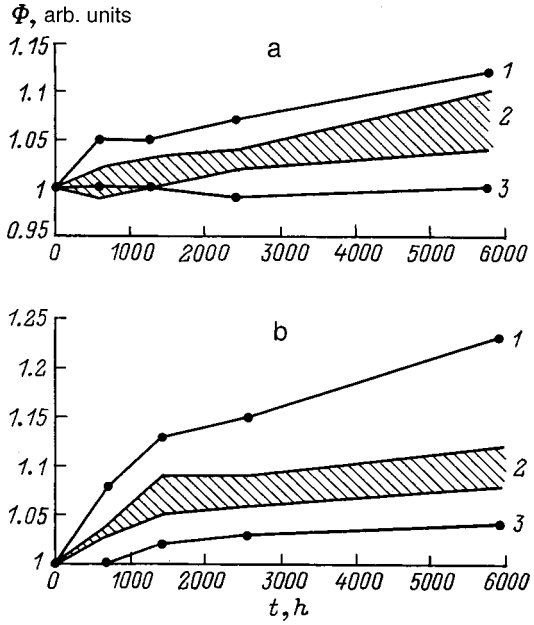

FIG. 2. Relative intensity of the radiation of breakdown LEDs versus the operating time: a — holder 1, b — holder 2; current through the sample, mA: *1* — 50, *2* — 10, *3* — 3; spectral range — integrated light with the FÉU-142 (*1*) and FÉU-84 (*2*,*3*) multipliers; interference light filters at wavelengths 361, 410, 474, 560, 608, and 687 nm (*2*).

the voltage with constant currents of 3, 10, and 50 mA remained unchanged during the operation to within three significant figures.

The current–voltage characteristic of the samples was measured before and after the tests. It was found that during prolonged operation the characteristic changed negligibly and touched the initial sections lying below the inflection. No clear connection between the changes in the current–voltage characteristic and the quantum yield was found.

The growth of the quantum yield of breakdown LEDs observed in this work during prolonged operation is opposite to the behavior of injection LEDs,[5] where the quantum yield drops because of the appearance of nonradiative recombination channels.

It seems that diffusion of charged impurities can occur under the influence of the strong electric field of the p–n junction. This diffusion expands the space-charge region and therefore increases the breakdown voltage and, correspondingly, changes the energy distribution of the high-energy carriers (it shifts the distribution into the region of high energies as a result of an increase in the drift length of free carriers in the strong-field region), which can increase the quantum yield of the breakdown electroluminescence and the fraction of high-energy photons in the spectrum. Near 2.6 eV, as well as in the red region of the spectrum, in alloyed silicon carbide p–n junctions injection electroluminescence bands due to radiative recombination of minority carriers are observed. Similar bands are present in the breakdown electroluminescence spectrum and can be explained[6] by radiative recombination of minority carriers arising in the neutral regions of the spectrum as a result of impact ionization by high-energy majority carriers which have left the strong-field region. During operation these bands can be quenched because of the

TABLE II.

Sample	Current, mA	U, V	$\gamma_U \times 10^3$, %/K	γ_{max}, %/K	γ_{min}, %/K	γ_{int}, %/K	$\eta_{int} \cdot 10^6$, photons per electron
164	50	12.61	−23	−.06	−.15	−.08	
	100	13.33	−15	−.02	−.08	−.04	.682
120	10	16.60	−3.0	−.02	−.1	−.03	
	50	17.24	0	+.01	−.05	−.01	2.85
	100	17.63	0.57	+.02	−.04	0	
11	10	18.50	−1.1	0	−.05	−.02	
	50	19.45	−1.0	+.04	−.02	+.01	2.12
	100	20.05	−1.0	+.06	0	+.03	
19–1	10	18.78	−3.2	0	−.03	−.02	
	50	19.47	−5.1	0	−.03	−.01	2.71
	100	20.01	−6.0	0	−.03	−.02	
90–1	50	20.28	−3.4	+.01	−.04	−.02	
*	50	20.28	−3.4	+.03	−.7	−.03	
14	10	20.78	1.9	0	−.04	−.02	
	50	21.50	0.93	+.01	−.03	−.01	2.29
	100	21.51	0.93	+.01	−.03	−.01	
18–1	10	21.45	−1.9	0	−.05	−.03	
	50	22.50	−3.6	+.01	−.03	−.02	3.45
	100	23.23	−2.6	+.01	−.03	−.01	
28–2	10	23.13	−3.5	0	−.05	−.02	
	50	25.09	−17	0	−.05	−.02	3.48
	80	26.05	−20	0	−.05	−.02	

appearance of nonradiative recombination channels, which is possible under conditions of a strong electric field and local heating in microplasmas. This quenching can partially compensate the growth of the main component of the radiation, due to transitions of high-energy carriers which occur without the participation of impurities and defects. These considerations can qualitatively explain the character of the curve in Fig. 1.

In choosing samples to investigate the temperature instability of the spectral distribution of the quantum yield, essentially the entire range of working voltages suitable for practical application of LEDs was covered. Certain characteristics of the samples and the measurement results are presented in Table II.

Where it is possible to do so with acceptable accuracy, the dependence $\gamma_{h\nu}(h\nu)$ was measured at the limits and at the center of the working range of currents in the stationary powering regime of a sample. Table II gives the maximum γ_{max} and minimum γ_{min} values of $\gamma_{h\nu}$ in the photon energy range 1.8–3.4 eV as well as the integrated value γ_{int} with respect to a nonselective photodetector, calculated for the same spectral range. For some samples the integrated internal quantum yield of the breakdown electroluminescence η_{int} in the photon energy range 1.8–3.4 eV is presented; it was estimated with an error not exceeding 100%.

For combined analysis, Table I also gives the voltage U on the sample (determined mainly by the uncompensated donor density in the initial crystals) at room temperature and the relative temperature coefficient γ_U of the voltage.

To determine the effect of self-absorption on the shape and temperature dependence of the spectral distribution of the quantum yield of breakdown LEDs, measurements were performed on the same typical sample with the radiation extracted from the p-region side as well as through the base crystal (this case is marked by an asterisk in Table II).

As one can see from Table II, the values of η_{int} in various samples differed negligibly, increasing appreciably with U. An exception is the low-voltage sample 164, whose operating regime seems to correspond to the tunneling breakdown region preceding the above-mentioned inflection in the current–voltage characteristic. Low values of η_{int} and a sharper dependence $\eta_{int}(U)$ are characteristic for such samples.

Uniquely low values of $\gamma_{h\nu}$ in the entire experimental spectral range, which, as a rule, do not exceed the limits (-0.05)–$(+0.05)$%/K, are observed when the radiation is extracted through a thin p region. A weak dependence of $\gamma_{h\nu}$ on U is observed. The low-voltage sample 164 is an exception. Such samples have high negative values of $\gamma_{h\nu}$, which increase in the direction of positive values with increasing U. Sample 11, which has high positive values of $\gamma_{h\nu}$, also stands out.

As the current through a sample increases, $\gamma_{h\nu}$, as a rule, shifts in the direction of positive values. This tendency appeared to a maximum degree for low-voltage samples as well as for sample No. 11. For emitters with working voltages above 20 V, the effect of a current in the range 10–100 mA on the temperature dependence of the quantum yield is negligible. It was established that a large increase in the modulus of negative values of $\gamma_{h\nu}$ starts when the current decreases to values close to the inflection in the current–voltage characteristic.

As one can see from Table II, a low negative or positive value of γ_U is characteristic for most samples. An exception is the low-voltage sample 164. Such samples possess a much sharper dependence $U(T)$, whose slope decreases as the

working regime approaches the inflection in the current–voltage characteristic and the transition to the section of sharp breakdown. The high-voltage sample 28–2 also stands out. For this sample high negative values of γ_U, which increase in modulus with increasing current, are observed. For this sample the influence of the serial resistance, which depends on the temperature more strongly than the resistance of the space-charge region, seems to be substantial. The high value of the differential resistance of sample 28–2 on the working current section agrees with this supposition.

The positive values of the temperature coefficient of the voltage which we observed in a number of alloyed p–n junctions based on silicon carbide prepared on crystal faces perpendicular to the crystallographic **C** axis in the present and previous works (for example, Ref. 1) do not agree with data on the ''anomalous'' temperature dependence of the electronic coefficient of impact ionization in α-SiC with electric field orientation $\mathbf{E} \| \mathbf{C}$.[7] The reason for this disagreement cannot yet be explained on the basis of existing models.[8]

The character of the distributions $\gamma_{h\nu}(h\nu)$ and $\xi(h\nu)$ is reproduced well for all samples. Figure 3 shows the spectra of a typical sample 90–1 with radiation extracted through a thin p region (curves *1*) as well as through the base crystal (curves *2*).

As one can see from the figure, the character of the curves $\gamma_{h\nu}(h\nu)$ and $\xi(h\nu)$ is similar in many respects. This suggests that one of the factors determining the temperature coefficient and its spectral dependence is the temperature shift of the main components of the radiation. The quantum yield of radiation increases on the approaching slope of the spectral band and decreases on the receding slope, and the relative change in quantum yield (i.e., the quantity $\gamma_{h\nu}$) in the absence of any effect from other factors should be proportional to the relative slope of the spectrum in this spectral range. As one can see from the figure, when radiation is extracted through the base SiC crystal, the temperature dependence of the optical absorption is the main factor determining the value and spectral behavior of the temperature coefficient in the region corresponding to the absorption edge (photon energies exceeding 2.6 eV). It also plays a large role in the impurity absorption range at the location of the well-known band with a maximum near 1.9 eV. As one can see from Table II, the integrated quantity γ_{int} improves for this sample by a factor of 1.5 when the radiation is extracted from the p-region side, primarily because any effect of the temperature shift of the optical absorption edge is eliminated.

The integrated temperature coefficient γ of the quantum yield of breakdown LEDs with respect to different types of photomultipliers is of special practical interest. A calculation for the most commonly used types of photocathodes, performed for sample 90–1 for radiation extracted through the p region, gives $\gamma = 0.108$–0.025%/K. For all types of photocathodes a gain by more than a factor of 2 obtains compared with radiation extraction through the base crystal (even though the region of the optical absorption edge contains a small fraction of the radiation of the LEDs, the maximum sensitivity of photocathodes lies in this region).

As one can see from Fig. 3, the influence of the temperature dependence of optical absorption also appears in the form of a dip near 3 eV in the curve $\gamma_{h\nu}(h\nu)$ when the radiation is extracted from p region side. The small fraction of radiation reflected from the crystal faces in the direction of the detecting system is subjected to self-absorption in this case.

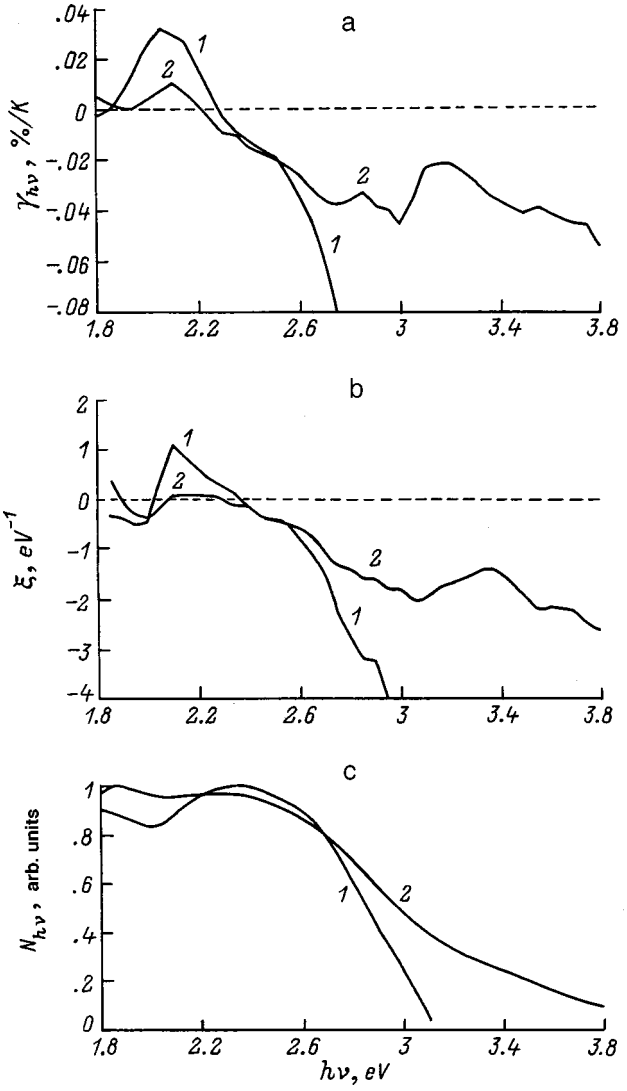

FIG. 3. Spectral dependences for a typical breakdown LED: a — distribution of the relative temperature coefficient of the spectral density of the quantum yield; b — distribution of the relative slope of the spectrum; c — photon energy distribution.

It has been established that, according to data for 40 emitters, γ for the UV component of the radiation (with respect to FEU-142) at a current of 100 mA lies in a comparatively narrow range from -0.15 to -0.06%/K. The high values of γ in the UV region compared with the visible region correlate with the larger slope of the emission spectrum, which makes it possible to attribute the indicated difference to the temperature shift.

In summary, we have determined the conditions under which LEDs, operating in the electric breakdown regime and possessing the maximum stability of the spectral composition of the radiation, are obtained using the alloying technology for obtaining p–n junctions on commercial silicon car-

bide crystals of the 6H polytype. Such devices can be used as standard sources of visible-range radiation. The preliminary data obtained on the parameters of LEDs in the UV region of the spectrum show that standard emitters can be produced for this range. More detailed investigations of the spectral and other characteristics of UV emitters are required.

We have shown that the quite high temperature stability, one of the most important requirements imposed on standard sources of radiation, can be achieved in the case of breakdown LEDs not by thermostating the device, as in the case of all conventional types of emitters, but rather as a result of the characteristic features of the mechanism of breakdown electroluminescence. This determines the possibility of producing small devices controlled by the powering current, which are ready for operation virtually instantaneously, and which can be used effectively in the pulsed regime.

The method proposed in the present paper for determining the temporal instability of sources of weak radiation in a wide spectral range is distinguished by high capacity, which makes it possible to use the method for metrological support of standard emitters under conditions of serial production.

[1] Yu. M. Altaĭskiĭ and A. M. Genkin, Zh. Tekh. Fiz. **52**, 543 (1982) [Sov. Phys. Tech. Phys. **27**, 350 (1982)].
[2] Yu. M. Altaĭskiĭ, A. M. Genkin, V. K. Genkina *et al.*, Élektronnaya tekhnika, Ser. 2, No. 4(10), 76 (1987).
[3] M. V. Belous, A. M. Genkin, V. K. Genkina *et al.*, Zh. Tekh. Fiz. **67**(1), 130 (1997) [Tech. Phys. **42**, 115 (1997)].
[4] A. M. Genkin, V. K. Genkina, and L. G. Ogneva, Diélektriki i Poluprovodniki, No. 32, 81 (1987).
[5] A. A. Ptashchenko, Zh. Prikl. Spektrosk. **33**, 781 (1980).
[6] L. A. Kosyachenko, A. V. Pivovar, and V. M. Sklyarchuk, Zh. Prikl. Spektrosk. **36**, 236 (1982).
[7] A. O. Konstantinov, Fiz. Tekh. Poluprovodn. **17**, 2124 (1983) [Sov. Phys. Semicond. **17**, 1358 (1983)].
[8] M. M. Anikin, M. E. Levinshteĭn, I. V. Popov *et al.*, Fiz. Tekh. Poluprovodn. **22**, 1574 (1988) [Sov. Phys. Semicond. **22**, 995 (1988)].

Translated by M. E. Alferieff

Model of current and optical conductivity switching in $YBa_2Cu_3O_{7-x}$ films

A. V. Mezenov

St. Petersburg State Electrical Engineering University, 197376 St. Petersburg, Russia

Ya. Vas'kevich

Bialystok Polytechnical Institute, Bialystok, Poland

(Submitted August 1, 1997; resubmitted January 11, 1999)
Zh. Tekh. Fiz. **69**, 77–82 (October 1999)

A model is proposed for the thermal and electrical responses of films of the high-T_c superconducting material $YBa_2Cu_3O_{7-x}$ to current and optical pulses. Numerical calculations are compared with experimental data for current pulses of duration 100 μs and laser pulses of duration 0.1 ns; this yields improved data on the thermal conductivity of thin $YBa_2Cu_3O_{7-x}$ films ($1.5-2$ W/m·K) and thermal resistance of the film–substrate contact (5×10^{-8} m^2·K/W) in the neighborhood of the superconducting transition. This model can be used for optimizing the film structure parameters and control regimes for switching elements for pulses lasting longer than 0.1 ns. © *1999 American Institute of Physics.* [S1063-7842(99)01110-1]

INTRODUCTION

Microwave cryoelectronic components for operation at liquid nitrogen temperatures, such as switches, modulators, clippers, etc., based on disrupting the superconducting state (S) of thin films of the high-T_c superconductor (HTSC) $YBa_2Cu_3O_{7-x}$ (YBCO) by supercritical current pulses or laser pulses, have been under intense development and study in recent years.[1,2] The main requirements for these components are speed and switching quality, which is evaluated in terms of the ratio of the resistances of the film in the resistive and superconducting states. The latter requires that the film undergo a transition into the normally conducting state (N). The physical limit on the rate at which a film can undergo a transition into the N state is determined by the electron–phonon interaction time, which is on the order of 1 ps.[3,4] The rapid transition of an entire microbridge film into the N state in the leading edge of a current pulse can be attained with high rates of rise of the current to average current densities on the order of $j=10^7$ A/cm^2 and with laser irradiation at irradiances on the order of 10^5 W/cm^2. The shortest experimentally observed resistive responses of HTSC microbridges with complete $S-N$ switching lasted less than 0.5 ns with an applied laser pulse of duration $\tau_l=0.1$ ns.[5] Responses with risetimes of less than 0.3 ns have been observed during switching with current pulses of duration $\tau_j=1.3$ ns.[6]

Experiments and calculations[7] show that the response of a film to this type of interaction with pulse lengths of more than 0.1 ns is determined mainly by heating of the film. The response has been simulated numerically since the characteristics of the electrical response of HTSC films to a controlling interaction depend strongly on the level of heating of the films and the thermophysical and geometric parameters of the film structure, and because of the difficulty of interpreting and quantitatively evaluating the experimental data. Numerical simulation makes it possible to determine the degree to which various factors affect the characteristics of the response and to establish the thermophysical parameters of HTSC films more accurately.

PHYSICAL JUSTIFICATION FOR THE MODEL

The computational models were constructed in conformity with the geometrical and temporal conditions typical of the switching components under consideration. At present, switching processes are usually studied using elements based on YBCO films. High speed can be achieved only with high quality films with a low temperature width of the superconducting transition ($\Delta T_c=0.5-1$ K). These films must be single crystalline with a crystallographic c axis perpendicular to the plane of the film. In order to improve the quality of films on some substrates, such as MgO, a buffer layer (sublayer) of material that is crystallographically matched to the structure of the HTSC film is first deposited. An HTSC film is produced in the form of a microbridge with the following characteristics: thickness $h_1=0.1-0.2$ μm, width $w=10-100$ μm, and length $l=0.1-15$ mm. The substrates usually have a thickness of $0.2-0.5$ mm and the sublayer thickness is $0.1-0.5$ μm.

An equilibrium state develops in several stages in an HTSC film. For example, under optical irradiation, the absorbed photons initially transfer energy to electrons at the penetration depth of the radiation, ~100 nm.[8] Electrons with energies on the order of 1 eV thermalize primarily through electron–electron collisions over times ~300 fs[9] and penetrate to an estimated[10,11] depth of $500-800$ nm at liquid nitrogen temperatures. As the excess energy of the electrons falls below 0.1 eV, thermalization through electron–phonon collisions begins to predominate[12] with a time constant of ~1 ps.[3,4] The phonon component of the thermal conductivity along the c axis in YBCO single crystals[13] is roughly two orders of magnitude greater than the electron component, i.e., heat is mainly carried by phonons whose mean free path at 90 K is on the order of 1 nm. These characteristic relax-

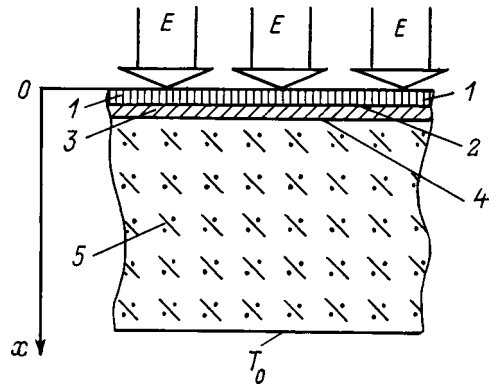

FIG. 1. Physical–geometrical model of an HTSC film element: *1* — HTSC film, *3* — sublayer, *5* — substrate, *2* and *4* — thermal contact regions; E is the irradiance of the radiation, I is the current, and T_0 is the thermostat temperature.

ation times suggest that a local thermodynamic state develops in YBCO films over times shorter than 10 ps.

Other mechanisms (besides bolometric) for conductivity switching in these control regimes ($\tau_l > 0.1$ ns; $h_l \geqslant 1000$ nm) either do not act at all, or end earlier (their effect can be characterized by the appearance of an effective resistance), or make a small contribution to the resulting response. Resistance owing to heating of the electrical subsystem (Testradi effect) can have an independent effect only in films with thicknesses $h_l < 30$ nm,[14] while in thicker films for times $\tau_l > 10$ ps, the effect reduces to heating of the film. An electrical response caused by the kinetic inductance of a microbridge has been analyzed in a number of papers[5,15] and it has been shown that it is also caused by rapid heating of the film, so it can be referred to as an inductive–bolometric effect. Estimates[6] show that the amplitude of this response is at least an order of magnitude smaller than the purely bolometric response.

Current disruption of the superconductivity of narrow films by currents much above critical also takes place over ~ 10 ps (on the order of the electromagnetic diffusion time)[2] and for modelling of the response to current pulses lasting > 0.1 ns, can be taken into account by introducing a resistance in the leading edge of the response pulse.

Given these factors, the basis of the mathematical model for the response of HTSC films to light and current pulses was assumed to be a thermal response mechanism, and the temperature distributions were calculated using the nonstationary heat conduction equations.[16]

MATHEMATICAL MODEL

Besides the superconductor (*1*), sublayer (*3*), and substrate (*5*) layers, the geometric model of an HTSC element shown in Fig. 1 includes layers (*2* and *4*) that model the region of thermal contacts between the HTSC film and the sublayer, as well as between the sublayer and the substrate. It is necessary to introduce contact layers because experimental data indicate that there is a large thermal resistance R_b in the contact regions. For example, the interface between an epitaxial YBCO film and an MgO (or $SrTiO_3$) substrate has been found[17] to have a thermal resistance $R_b \approx 10^{-8}$ $m^2 \cdot K/W$ comparable to the thermal resistance of an HTSC film of thickness $100-200$ nm. The film l is acted on by optical radiation with an irradiance E and a transport current I flows along the film.

For the characteristic geometric dimensions of microbridges, when the length and width of the HTSC film are at least 50 times its thickness, in the calculations of pulsed heating of the film a one-dimensional model can be used with the assumption that heat propagates along the x axis perpendicular to the plane of the HTSC film. The system of time dependent heat conduction equations for the layers of a film structure have the following general form:

$$c_k \rho_k \frac{\partial T}{\partial t} = \lambda_k \frac{\partial^2 T}{\partial x^2} + p_k(x,t,T), \tag{1}$$

where k is the number of the layer of the film structure (Fig. 1); T is the temperature; t is the time; x is the coordinate; c_k, ρ_k, and λ_k are the specific heat, density, and thermal conductivity, respectively, of the material in the kth layer; and, p is the volume power density of heat released in the plane at position x.

Since dielectric, optically transparent substrates and sublayers are usually used in film structures, we may assume that heat is released only in the superconducting layer and that Eqs. (1) will be homogeneous ($p=0$) for layers 2–5.

The thermophysical parameters λ_k, c_k, and ρ_k were assumed to be constant, since in the temperature range from 77 K (liquid nitrogen) to 90 K (the onset of the superconducting transition) of greatest interest for analyzing the response characteristics, the errors associated with neglecting the temperature dependences of these parameters is less than 15%.

The volume power of the heat source in the HTSC layer, p, is made up, in general, of the power p_l generated by the absorption of the light and by Joule heating, p_j, and depends on the time, position, and temperature of the superconductor, i.e.,

$$p(x,t,T) = p_l(x,t) + p_j(t,T).$$

Since the penetration depth of visible and near infrared light in YBCO, ~ 100 nm,[8] is commensurate with the thickness of the HTSC film, on taking the transmission of the film into account, the radiative power q absorbed per unit area of the film is given by

$$q(t) = q_0(t)(1-R_\nu)(1-e^{-\kappa h_1}),$$

where R_ν is the spectral reflectivity of the film, κ is the absorption coefficient of the film material, and $q_0(t)$ is the surface flux of the incident radiation.

Since the diffusion mean free path for fast electrons exceeds the HTSC layer thickness, we may assume that on a time scale $\tau > 100$ ps, the heat sources are uniformly distributed over the film thickness, i.e.,

$$p_l(x,t) = \mathrm{const}_x(t) = \frac{q(t)}{h_l}.$$

The time variation of the laser pulses was approximated by a rectangular or gaussian shape. The shape of pico- and

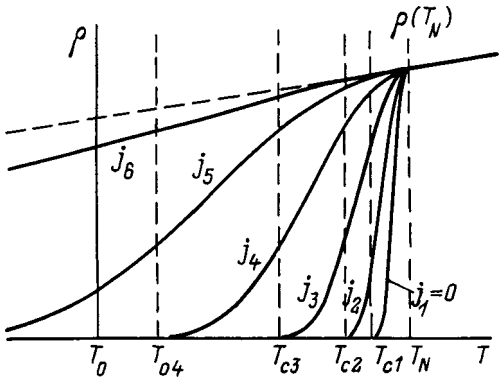

FIG. 2. Approximating the $\rho(T,j)$ curve by piecewise continuous functions.

nanosecond light pulses is usually assumed to the rectangular, since in this case the response is determined mainly by the total energy delivered to the film. The amplitude of the pulse was then determined from the density of the total energy of the optical signal, $q_0=E/\tau_l$.

The volume Joule heating power p_j is determined assuming a uniform distribution of the current over the film cross section, which is valid after the disruption of the S state, i.e.,

$$p_j(t,T)=j^2(t)\cdot\rho(T,j), \tag{2}$$

where j is the current density and ρ is the specific electrical resistivity of the HTSC film.

Modelling the current interaction is a complex task, because in the region of the phase transition ρ depends on the current density,[5] as well as on the temperature. The following algorithm was carried out in the model. The onset temperature of the $N-S$ transition, T_N, and $\rho(T_N)$ are assumed constant. The $j_c(T)$ curve, which is linear over a wide temperature range near the superconducting transition, is essentially used to solve the inverse problem of determining the transition temperature T_c from the current density in the film. The $\rho(T)$ curve between these points was approximated by different functions selected by analyzing the experimental pulsed current–voltage characteristics of the HTSC films. The function $\rho(T,j)$ used in this paper, which contains[18] the square of the sine, is plotted in Fig. 2. The critical current densities at the temperatures of liquid nitrogen and the superconducting transition were used as fit parameters for comparing the calculated and experimental response characteristics.

Therefore, the effect of the transport current was taken into account in the model, on one hand, through its effect on T_c and ρ at a given temperature and, on the other, through the heat release (2). In the front of the current pulse a transition from one $\rho(T)$ curve to the other at a fixed temperature is carried out in the program (Fig. 2); this is fully justified, since the time to disrupt the superconducting state with a current is ~ 10 ps, an order of magnitude shorter than the shortest pulsed interactions considered here. The resulting value of ρ is used to calculate the resistance of the film assuming a uniform distribution of resistivity. This modelling approach is suitable for analyzing the responses to current pulses well above critical, sufficient to bring all or most of the length of the microbridge into a normally conducting state in the leading edge of the current pulse. A two-dimensional model is used to model the nonuniform disruption of the superconducting state along a bridge; in that model the length of the thermal domain formed during the current pulse risetime and its resistance are determined using dependences of the film resistance in the leading edge of the current pulse, $R(j,T)$, that are analogous to the $\rho(j,T)$ curves shown in Fig. 2.[19]

In the complete heat conduction boundary value problem, Eq. (1) for the five layers is supplemented by boundary and initial conditions. At low temperatures, heat removal from the outer surface of the film into the surrounding medium is negligible and for $x=0$ we can assume Neumann (type II) boundary conditions,

$$\left.\frac{\partial T(t)}{\partial x}\right|_{x=0^+}=0.$$

Type IV boundary conditions hold for the inner interfaces of the layers, x_k; these represent equality of the fluxes on the left and right,

$$\lambda_k\left.\frac{\partial T(t)}{\partial x}\right|_{x_k^-}=\lambda_{k+1}\left.\frac{\partial T(t)}{\partial x}\right|_{x_k^+},$$

where $k=1-(N-1)$ is the number of a layer from the HTSC film to the layer adjacent to the substrate (N is the number of layers), while

$$x_k=\sum_{k=1}^{N-1} h_k$$

represents the coordinate of the interface of the layers, where h_k is the thickness of the layers.

For the substrate surface x_5, which adjoins the heat sink, a Dirichlet (type I) condition is specified: $T(x_5,t)=T_0$. The

TABLE I.

h_1, μm	w, μm	l, mm	T_N, K	T_c, K	$\rho(T_N)$, m$\Omega\cdot$cm	j_c (77 K), A/cm^2	R_ν,	κ, μm^{-1}
			$j=0$					
0.1	100	1	90	86.7	0.1	10^6	0.1	10

λ_1, W/m$\cdot$K	c_1, J/kgK	ρ_1, kg/m^3	R_b, m^{2}K/W	λ_3, W/m$\cdot$K	c_3, J/kgK	ρ_3, kg/m^3	T_0, K	τ_j, s	I, mA
5	150	6350	10^{-8}	5	200	5000	86	10^{-4}	20–40

initial condition $T(x,0)=T_0$ reflects the fact that at time $t=0$ the entire structure is at the temperature of the thermostat.

The heat conduction boundary value problem formulated here is a system of homogeneous, inhomogeneous, and nonstationary differential equations that are linear in the thermophysical properties of the materials and boundary conditions, but have a highly nonlinear volume heat source p_j in the first layer. The problem can only be solved numerically, with the entire set of parameters of the structure, material, and external interactions. The method chosen was a finite difference method employing a purely implicit Euler scheme. The system of algebraic equations was linearized by Newton's method. This yields a system of equations with a tridiagonal matrix which is rapidly solved by iteration.

In the program for executing the model, the parameters of the electrical circuit for hookup of the HTSC film are also established.

TEST OF THE MODEL

The model was tested by comparing experimental responses to rectangular current pulses[20] of duration $\tau_j=100\,\mu$s and to laser pulses[5] of duration $\tau_l=0.1$ ns with calculations for these conditions.

Thin layers of YBCO deposited on $SrTiO_3$ substrates were studied in Ref. 20. The parameters of the structure and external interaction employed in the calculations are listed in Table I. Judging from the critical temperatures T_c and T_N, the critical current density j_c (77 K), and the specific resistance $\rho(T_N)$, the films used in the experiments were of good quality.

The experimental responses of HTSC microbridges (in the form of voltage pulses) to 100 μs current pulses with different amplitudes I are shown in Fig. 3a. The responses calculated for conditions corresponding to these experiments (Table I) are shown in Fig. 3b. The calculated and experimental curves are in good qualitative and quantitative agreement. In these curves, a sharp rise can be seen clearly during the thermally unstable conditions that occur when the Joule heating released in the film exceeds the amount of heat removed from the microbridge into the substrate.

The calculated responses of HTSC film elements to optical pulses were compared with experimental data from Ref. 5. That paper is of interest because the responses obtained there had good time resolution and the applied laser pulse was very short. In addition, the properties of the $S-N$ transition in the experimental samples of $YBa_2Cu_3O_{7-x}$ films deposited on $LaAlO_3$ substrates were studied quite thoroughly.

In the experiments, the HTSC films were irradiated by pulses of the second harmonic of a YAG:Nd laser (532 nm) with a half-width $\tau_l=0.1$ ns and a surface energy density $E=10^{-4}$ J/cm^2. A constant current, slightly above critical, was fed through the film to raise the initial temperature T_0 of the HTSC films by ΔT_c into the onset region for the resistive transition. The parameters characterizing the experimental conditions and assumed as initial parameters for the calculations are listed in Table II. The values of λ_1, c_1, ρ_1, R_ν,

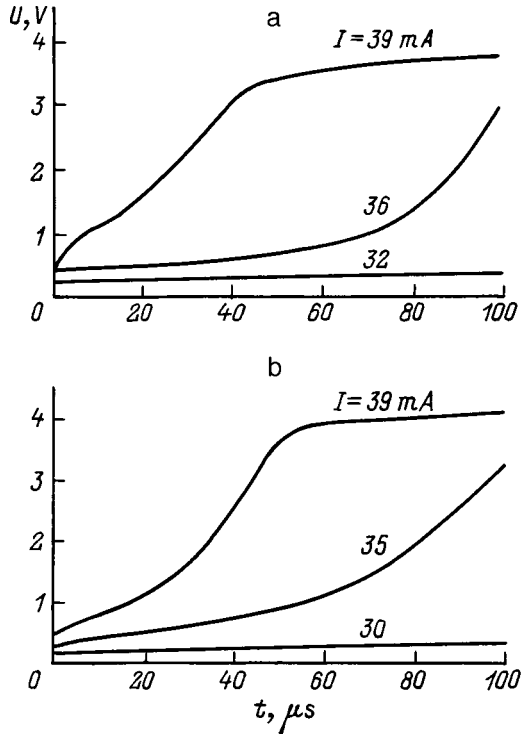

FIG. 3. Responses of a $YBa_2Cu_3O_{7-x}$ microbridge to current pulses of duration $\tau_j=100\,\mu$s with different amplitudes I: a — experiment[20], b — calculations.

and κ are the same as in Table I. The following parameters were used for the $LaAlO_3$ substrate: $\lambda=50$ W/m·K, $c=100$ J/kg·K, and $\rho=6000$ kg/m^3.

Figure 4a shows the experimental responses of films of different thicknesses for the currents I listed in Table II. The duration of the response (at half maximum) decreased with film thickness from 8 ns for $h_1=0.22\,\mu$m to ~ 1 ns for $h_1=0.03\,\mu$m. The calculated responses are shown in Fig. 4b. The general variations in the theoretical and experimental curves are in good agreement. This confirms the validity of the model choice of the bolometric mechanism as the major mechanism for conductivity switching in HTSC films as they are brought into the resistive state. The experimental results were also interpreted in terms of a thermal response in Ref. 5, but a simpler heat conduction model was used there.

The thermal conductivity of the HTSC film, λ_1, and the thermal resistance of the film–substrate contact, R_b, were chosen as fitting parameters in the simulations. The best agreement between the calculated and experimental responses was obtained with thermal conductivities of $YBa_2Cu_3O_{7-x}$ in the range 1.5–2 W/m·K and R_b equal to $\sim 5\times 10^{-8}$ m^2·K/W. This value of the film thermal conductivity is close to that of single crystals in the c direction, but is substantially higher than that given in Ref. 17. The contact thermal resistance was modelled, as in Ref. 17, by a layer of thickness $h_1=1$ nm (which is comparable to the lattice constant of $YBa_2Cu_3O_{7-x}$ along the axis of orientation of the film, 1.16 nm) with a thermal conductivity of ~ 0.02 W/m·K; the value of R_b was roughly a factor of 2 smaller than that given in Ref. 21 for a $YBa_2Cu_3O_{7-x}/LaAlO_3$ structure.

TABLE II.

h_1, μm	w, μm	l, mm	T_N, K ($j=0$)	T_c, K ($j=0$)	$\rho(T_N)$, m$\Omega\cdot$cm	j_c (77 K), A/cm^2	T_0, K	I, mA	λ_l, W/m$\cdot$K	R_b, m$^2\cdot$K/W
0.03	20	0.2	89	86	0.18	8.3×10^5	73	20	2.0	5×10^{-8}
0.13	20	0.2	89	87	0.16	9.6×10^5	70	13	1.8	5×10^{-8}
0.22	20	0.2	91	90	0.11	2.3×10^6	84	80	2.0	5×10^{-8}

CONCLUSIONS

The proposed physical and mathematical model for the thermal and electrical response of HTSC films, which takes into account the disruption of superconductivity by a current and the heating of the film by Joule heating and optical radiation, adequately reflects the response to pulsed current and optical interactions lasting longer than 0.1 ns. We have demonstrated the possibility of determining the thermal conductivity of thin $YBa_2Cu_3O_{7-x}$ films and the thermal resistance of film–substrate contacts in the neighborhood of the superconducting transition temperature by comparing the experimental responses to short optical pulses with model calculations of the electrical responses of HTSC elements.

Our mathematical model of the response of HTSC elements to pulsed currents and optical radiation can be used for optimizing the parameters of the driver signals and the geometric parameters of film structures in order to minimize the recovery time for the superconducting state after termination of the driver signal.

This work was supported by the Polish Committee on Scientific Research (Grants PB-W/WE/9/97 and KBN/21/3/98) and the Federal Applied Project on Integrated Circuits (Project No. 141) and Grant No. 96-15-98271.

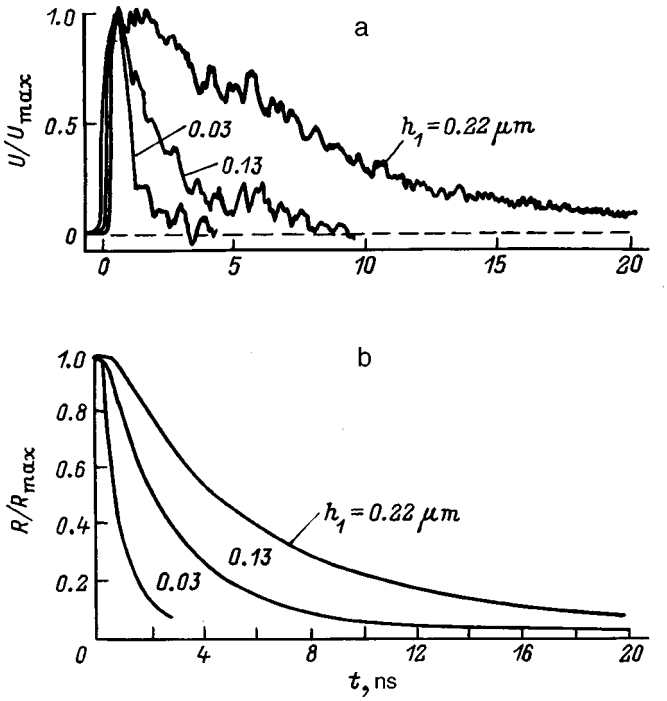

FIG. 4. Responses of $YBa_2Cu_3O_{7-x}$ films of different thicknesses h_1 to optical pulses: a — experiment,[5] b — calculations.

[1] O. G. Vendik, M. M. Gaĭdukov, A. B. Kozyrev *et al.*, Obzory po VTSP, No. 4(8), pp. 43–100 (1992).
[2] A. B. Kozyrev, T. B. Samoĭlova, and S. Yu. Shaferova, Sverkhprovodimost' (KIAE) **6**, 823 (1993).
[3] M. E. Gershenzon, V. V. Golovlev, I. B. Kedich *et al.*, JETP Lett. **52**, 602 (1990).
[4] J. M. Chwalek, C. Uher, J. F. Whitaker *et al.*, Appl. Phys. Lett. **57**, 1696 (1990).
[5] F. A. Hegmann and J. S. Preston, Phys. Rev. B **48**, 16023 (1993).
[6] B. S. Karasik, M. A. Zorin, I. I. Milostnaya *et al.*, J. Appl. Phys. **77**, 4074 (1995).
[7] H. Yasuoka, H. Mazaki, T. Terashima *et al.*, Physica C **175**, 192 (1991).
[8] Ya. Vas'kevich and A. V. Mezenov, Izv. GETU (LÉTI), No. 481, pp. 42–49 (1995).
[9] S. G. Han, Z. V. Vardeny, O. G. Symko *et al.*, IEEE Trans. Magn. **MAG-27**, 1548 (1991).
[10] E. M. Gershenzon, G. N. Gol'tsman, A. V. Sergeev *et al.*, in *Proceedings of the European Conference on High-T_c Thin Films and Single Crystals*, Ustron (1989), pp. 457–462.
[11] P. H. Ballentine, A. M. Kadin, W. R. Donaldson *et al.*, Proc. Soc. Photo-Opt. Instrum. Eng. (SPIE), pp. 134–143 (1990).
[12] N. Bluzer, J. Appl. Phys. **71**, 1336 (1992).
[13] S. J. Hagen, Z. Z. Wang, and N. P. Ong, Phys. Rev. B **40**, 9389 (1989).
[14] M. E. Gershenzon, V. V. Golovlev, I. B. Kedich *et al.*, JETP Lett. **52**, 602 (1990).
[15] D. Ladret, B. Cabon, T. V. Dinh *et al.*, IEEE Trans. Appl. Supercond. **AS-5**, 3482 (1995).
[16] B. A. Grigor'ev, *Pulsed Heating by Radiation. Part 2. Nonstationary Temperature Fields under Pulsed Radiant Heating* [in Russian], Nauka, Moscow (1974).
[17] C. D. Marshall, A. Tokmakoff, and I. M. Fishman, J. Appl. Phys. **73**, 850 (1993).
[18] Ya. Vas'kevich, A. V. Mezenov, and A. V. Yankevich, Izv. GÉTU (LÉTI), No. 506, pp. 41–47 (1995).
[19] A. V. Mezenov and A. V. Yankevich, Izv. GÉTU (LÉTI), No. 506, pp. 37–41 (1997).
[20] Yu. D. Varlamov, A. N. Volkov, M. R. Predtechenskiĭ *et al.*, Sverkhprovodimost' (KIAE) **5**, 1736 (1992) [Sverkhprovodimost: Fiz., Khim., Tekh. **5**, 1672 (1992)].
[21] M. Nahum, S. Verghese, P. L. Richards *et al.*, Appl. Phys. Lett. **59**, 2034 (1991).

Translated by D. H. McNeill

Ion surface cyclotron waves at plasma–metal interfaces

V. O. Girka and I. V. Pavlenko

Kharkov State University, 310077 Kharkov, Ukraine

V. I. Lapshin

Kharkov Physicotechnical Institute National Science Center, National Academy of Sciences of Ukraine, 310108 Kharkov, Ukraine

(Submitted February 28, 1998; resubmitted March 2, 1999)
Zh. Tekh. Fiz. **69**, 83–89 (October 1999)

The dispersion properties of slow electromagnetic surface waves propagating across a constant external magnetic field and along a plane plasma–metal interface at harmonics of the ion cyclotron frequency are studied. The motion of the plasma particles is described by a Vlasov–Boltzmann kinetic equation. The effects of the plasma size, the dielectric permittivity of the transition region between the plasma and metal, and the magnitude of the constant external magnetic field on the dispersion characteristics of ion surface cyclotron waves are studied. © *1999 American Institute of Physics.* [S1063-7842(99)01210-6]

INTRODUCTION

The dispersion properties and propagation behavior of volume cyclotron waves have been extensively studied.[1–5] These waves are widely used in such areas of science and technology as controlled thermonuclear fusion, plasma physics, and electronics. Meanwhile, the existence of plasmas within limited volumes in the laboratory means that surface waves may propagate in these plasmas. The dispersion properties of surface waves and numerous effects associated with their propagation have been under intense theoretical and experimental study.[6–9] Their integral properties are determined by the parameters of the plasma device, such as the plasma density and temperature, the direction and magnitude of the external magnetic field, and the material and shape of the surface surrounding the plasma. In controlled thermonuclear fusion experiments, volume cyclotron waves are excited by external rf energy sources and are used for auxiliary plasma heating and the generation of noninductive currents in tokamaks.[10–13] The conditions under which they are excited can also favor the excitation of surface cyclotron waves. However, since the energy of a surface wave is localized along the plasma boundary, the existence of wave excitations of this type may cause undesirable heating of the plasma edge and, therefore, enhance the interaction of plasma particles with the structural material of the machine. Increasing the energy of plasma particles in the divertor region of thermonuclear machines is particular undesirable, since this may raise the flux of charged particles to the walls.

In recent years, surface waves have come into widespread use in plasma technologies for precision working of materials, thin film deposition, the creation of plasma sources, etc.[14,15] Plasma electronics is yet another domain of application of surface waves. The possibility of using the distinctive features of surface waves for developing the physical foundations of the operation of various rf electronic devices has been discussed.[16–18]

Surface electromagnetic perturbations at harmonics of the cyclotron frequencies at the boundary between plasmas and metals coated with a protective layer of insulator have been studied in the case where a constant external magnetic field is oriented perpendicular to the interface.[19,20] This orientation of the vector $\mathbf{B}_0$ can be realized in the divertor region of thermonuclear machines, as well as when a ferromagnetic metal is used as the wall of a waveguide structure. The spatial distribution of the field of rf surface electron cyclotron waves has been studied,[19,20] including the dependence of their frequency on the magnitude of the external magnetic field, on the characteristics of the insulator coating separating the plasma from the metal,[19] and on the inhomogeneity of the plasma density.[20] It has also been shown that the penetration depth of these waves into the depth of the plasma is much greater than the wavelength of surface electron cyclotron waves and that the characteristics of these waves depend on the parameters of a insulator layer separating the plasma from a metal wall. However, the dependence of the properties of surface cyclotron waves on the parameters of a waveguide structure remains an open question in the case where the penetration depth of the field into the plasma for these waves is of the same order as or less than the size of the plasma region. This paper is devoted to an investigation of this question.

1. DESCRIPTION OF MODEL AND BASIC EQUATIONS

Let us consider a uniform plasma layer occupying the space $0<z<a_{pl}$. The ideally conducting metallic walls of a waveguide lie in the regions $z<-a_d$ and $z>a$, where $a=a_{pl}+a_d$. The waveguide walls are coated with an insulator layer of thickness a_d. It occupies a region whose dimensions are determined by the inequalities $a_{pl}<z<a$ and $-a_d<z<0$. The dielectric permittivity of the insulator is

ε_d. The plasma–insulator interface is assumed to be sharp. A constant external magnetic field $\mathbf{B}_0 \| \mathbf{Z}$ is applied perpendicular to the plasma surface.

A hydrodynamic description of the motion of the plasma particles does not account correctly for the finiteness of the Larmor radius, so we shall employ a Vlasov–Boltzmann kinetic equation for that purpose. The distribution function for the particles in the unperturbed plasma is assumed to be maxwellian. We shall consider a mirror model for the reflection of particles from the plasma–insulator interface. This approach is widely used for problems of this type[8] and permits fairly simple calculations of the wave damping associated with the interaction between the plasma particles and the surface surrounding the plasma. The space is assumed to be uniform along the **y** direction and the dispersion of the medium along the direction of the external magnetic field is assumed to be weak, i.e., $k_3 v_{T\alpha} \ll (\omega - s|\omega_\alpha|)$, where k_3 is the z component of the wave vector of the surface cyclotron wave, $v_{T\alpha}$ and ω_α are, respectively, the thermal speed and cyclotron frequency of the particles of species α, with $\alpha = e$ for the electrons and $\alpha = i$ for the ions, and s is an integer. This inequality means that the penetration depth of the surface cyclotron waves into the plasma is much greater than the Larmor radius $\rho_\alpha = v_{T\alpha}/|\omega_\alpha|$ of the plasma particles. The expressions for the components of the dielectric permittivity tensor of the plasma, ε_{ik}, obtained under these assumptions are the same as those for an infinite magnetoactive plasma.[3–5]

The electromagnetic field of a surface cyclotron wave obeys Maxwell's equations. The x dependence of the field of a surface cyclotron wave was chosen to have the form $E, H \propto \exp(ik_1 x - i\omega t)$, where k_1 and ω are the component of the wave vector along the propagation direction and the wave frequency, respectively. Assuming that space is uniform along the **y** direction, i.e., $(\partial/\partial y) = 0$, and that the electromagnetic perturbations are slow ($|\omega^2 \varepsilon_{ik}| \ll k_1^2 c^2$, where ε_{ik} is the dielectric tensor of a magnetoactive plasma), Maxwell's equations break down into two subsystems. One, with the electromagnetic field components E_x, H_y, and E_z, describes the slow surface cyclotron wave. In order to solve Maxwell's equations, we expand the components of the surface cyclotron wave field in a Fourier series. The system of differential equations transforms to a system of algebraic equations for the Fourier coefficients of the wave field in the plasma region,

$$ik\varepsilon_{11}E_1 - i\alpha_n H_2 + ik\varepsilon_{13}E_3 = \frac{(-1)^n H_y(a_{pl}) - H_y(0)}{2a_{pl}},$$

$$i\alpha_n E_1 - ikH_z - ik_1 E_3 = -\frac{(-1)^n E_x(a_{pl}) - E_x(0)}{2a_{pl}},$$

$$k\varepsilon_{31}E_1 + k_1 H_2 + k\varepsilon_{33}E_3 = 0, \qquad (1)$$

where $k = \omega/c$ and $\alpha_n = \pi n/a_{pl}$ is the variable of the Fourier sum.

It has been shown[8] that the damping of surface cyclotron waves owing to the dispersion of the medium is weak. Since we shall not be concerned with its magnitude here, we shall ignore the imaginary terms in the plasma dielectric tensor ε_{ik}. The real part of ε_{ik} determines the dispersion of the surface cyclotron waves. Here we give expressions for the terms of the tensor which enter in the system of Eqs. (1):

$$\varepsilon_{11} = 1 - \sum_\alpha \sum_{n=-\infty}^{\infty} \frac{n^2 \omega_{p\alpha}^2}{y_\alpha \omega(\omega - n\omega_\alpha)} I_n(y_\alpha)\exp(-y_\alpha),$$

$$\varepsilon_{13} = \varepsilon_{31} = -\sum_\alpha \sum_{n=-\infty}^{\infty} \frac{s\omega_{p\alpha}^2 \omega_\alpha k_3}{\omega(\omega - n\omega_\alpha)^2 k_1} I_n(y_\alpha)\exp(-y_\alpha),$$

$$\varepsilon_{33} = 1 - \sum_\alpha \sum_{n=-\infty}^{\infty} \frac{\omega_{p\alpha}^2}{(\omega(\omega - n\omega_\alpha)} I_n(y_\alpha)\exp(-y_\alpha), \qquad (2)$$

where $\omega_{p\alpha}$ is the plasma frequency, $I_n(x)$ is the modified Bessel function,[21] and $y_\alpha = k_1^2 \rho_\alpha^2/2$.

We shall solve the system of Eqs. (1) for the Fourier component E_1.

In order to determine the tangential components of the electric and magnetic fields of the surface cyclotron waves at the boundaries of the plasma layer (E_x, H_y) using Watson's method, we calculate the corresponding Fourier sums. The system of Eqs. (1) can also be used to find the structure of the surface cyclotron wave field in the insulator. For this purpose, it is necessary to made the following substitutions: $\varepsilon_{11} \to \varepsilon_d$, $\varepsilon_{13} = \varepsilon_{31} \to 0$, and $\varepsilon_{33} \to \varepsilon_d$, $a_{pl} \to a_d$.

2. DISPERSION RELATION

Once an expression has been obtained for the tangential components of the electric and magnetic fields of the surface cyclotron wave at the boundaries of the plasma and the two insulator layers, it is possible to calculate the corresponding impedance ratios ($Z = E_x/iH_y$),

$$Z_{pl}(0) = C1\left\{\frac{Z_{pl}(a_{pl})\cosh(|k_1|ba_{pl}) + C1\sinh(|k_1|ba_{pl})}{Z_{pl}(a_{pl})\sinh(|k_1|ba_{pl}) + C1\cosh(|k_1|ba_{pl})}\right\},$$

$$Z_d(-a_d) = C2\left\{\frac{Z_a(0)\cosh(|k_1|a_d) + C2\sinh(|k_1|a_d)}{Z_d(0)\sinh(|k_1|a_d) + C2\cosh(|k_1|a_d)}\right\},$$

$$Z_d(a_{pl}) = C2\left\{\frac{Z_d(a)\cosh(|k_1|a_d) + C2\sinh(|k_1|a_d)}{Z_d(a)\sinh(|k_1|a_d) + C2\cosh(|k_1|a_d)}\right\}, \qquad (3)$$

where

$$C1 = \frac{|k_1|}{k\sqrt{\varepsilon_{11}(\varepsilon_{33} + 2B)}}, \quad C2 = \frac{|k_1|}{k\varepsilon_d},$$

$$b = \sqrt{\frac{\varepsilon_{11}}{\varepsilon_{33} + 2B}}, \quad B = \frac{k_1}{k_3}\varepsilon_{13}.$$

B is independent of the z component of wave vector k_3 of the surface cyclotron wave. We obtain a dispersion relation for the surface cyclotron wave on equating the corresponding surface impedances at the plasma–insulator boundaries $Z_{pl}(0) = Z_a(0)$ and $Z_{pl}(a_{pl}) = Z_d(a_p)$. Because of the infinite conductivity of the metal, the surface impedances of the insulator at the interfaces with the metal must go to zero, i.e., $Z_a(-a_d) = 0$ and $Z_d(a) = 0$. The dispersion relation for this waveguide structure has the form

$$\sqrt{\varepsilon_{11}(\varepsilon_{33}+2B)}\tanh(|k_1|a_d)\tanh(|k_1|ba_{pl})=\varepsilon_d. \quad (4)$$

The effect of the plasma density inhomogeneity on the dispersion of surface waves is often studied by modelling the nonuniformity of the plasma profile by a set of uniform layers with different thicknesses and different densities.[7,8,18] In this case, the surface cyclotron wave dispersion relation will be different from Eq. (4). We shall write it for the case in which the plasma inhomogeneity is modelled by a set of N layers with thicknesses $a_1, a_2, \ldots, a_N$ and corresponding densities $n_1, n_2, \ldots, n_N$, i.e.,

$$\sqrt{\varepsilon_{11}(n_0)(\varepsilon_{33}(n_0)+2B(n_0))}=\sum_{i=1}^{N} f(n_1) + \varepsilon_d\coth(|k_1|a_d), \quad (5)$$

where $f(n_i)=\varepsilon_{11}(n_i)(\varepsilon_{33}(n_i)+2B(n_i))\tanh(|k_1|b(n_i)a_1)$.

Equation (5) can be studied for arbitrary values of the parameters only by numerical methods. The results of such a study are given in Sec. 4.

3. DISPERSION PROPERTIES OF SURFACE CYCLOTRON WAVES

An analysis of Eq. (4) shows that the dispersion of surface cyclotron waves depends substantially on the transverse size of the plasma layer. In the limit of a thick plasma layer ($|k_1|b\gg 1$), the dispersion relation (4) transforms into the dispersion relations for surface cyclotron waves in the well-studied limits of thick ($|k_1|a_d\gg 1$)[8] and thin ($|k_1|a_d\ll 1$)[19] insulator layers. In the limit of a thin plasma layer ($|k_1|b \ll 1$), the dispersion relation (4) transforms to the simpler form

$$|\varepsilon_1|\tanh(|k_1|a_d)|k_1|a_{pl}=2\varepsilon_d. \quad (6)$$

Simple analytic solutions of Eq. (6) in the form

$$\omega=\frac{s|\omega_i|}{1-h_i}: \quad |h_i|\ll 1$$

can be obtained only in the limit of thick ($|k_1|a_d\gg 1$) and thin ($|k_1|a_d\ll 1$) insulator layers.

For thin plasma and thick insulator layers, in the long ($y_i\ll 1$) and short ($y_i\gg 1$) wavelength limits, the solution of the dispersion relation (6) has the form

$$h_i=\left(\frac{\omega_{pi}}{\omega_i}\right)^2\frac{s^2y_i^{s-1}}{2^{s+1}s!}k_{pl}, \quad y_i\ll 1,$$

$$h_i=\left(\frac{\omega_{pi}}{\omega_i}\right)^2\frac{s^2}{\sqrt{\pi}(2y_i)^{3/2}}k_{pl}, \quad y_i\gg\left(\frac{\omega_{pi}}{\omega_i}\right)^{4/3}, \quad (7)$$

where $k_{pl}=(k_1a_{pl})/\varepsilon_d$.

For thin plasma and thin insulator layers, we have

$$h_i=\left(\frac{\omega_{pi}}{\omega_i}\right)^2\frac{s^2y_i^{s-1}}{2^{s+1}s!}k_{pl}k_d\varepsilon_d, \quad y_i\ll 1,$$

$$h_i=\left(\frac{\omega_{pi}}{\omega_i}\right)^2\frac{s^2}{\sqrt{\pi}(2y_i)^{3/2}}k_{pl}k_d\varepsilon_d, \quad y_i\gg\left(\frac{\omega_{pi}}{\omega_i}\right)^{4/3}. \quad (8)$$

Equations (7) and (8) were obtained first. They imply that the dispersion of ion surface cyclotron waves depends significantly on the size of the plasma layer. A gradual reduction in k_{pl} leads initially to a drop in the frequency of the surface cyclotron waves and in their group velocity and, finally, to the disappearance of surface cyclotron waves as an eigenmode of the waveguide structure under consideration.

An analysis of the asymptotic solutions (7) and (8) shows that as the magnitude of the constant external magnetic field $\mathbf{B}_0$ is reduced, the region in which the surface cyclotron waves exist shifts toward longer wavelength perturbations in the limit $y_i\ll 1$ and to shorter wavelength perturbations in the limit $y_i\gg 1$. The conditions for existence of surface cyclotron waves deteriorate (collisional damping increases[8]) as $a_d\to 0$ and, in the absence of an insulator layer, surface cyclotron waves do not propagate.

4. NUMERICAL RESULTS

The dispersion relation (4) was solved numerically for arbitrary thicknesses of the plasma and insulator layers. This numerical study of the dispersion of surface cyclotron waves was done for the first four harmonics of the ion cyclotron waves. The dispersion of surface cyclotron waves in the waveguide structure considered here differs from that in the semi-infinite plasma–semi-infinite insulator system considered in detail previously.[8] At long perturbation wavelengths, there is a change in the dispersion of surface cyclotron waves. A segment with forward dispersion (the group and phase velocities of the waves have the same direction) develops. This occurs when, as the thickness of the insulator region or the size of the plasma region are reduced, the energy flux of the wave in the insulator region becomes comparable to that in the plasma region. As the plasma region is reduced in size, this equality is achieved when the penetration depth of the surface cyclotron wave field into the plasma,

$$\lambda_{pl}=\frac{2\pi}{k_1}\sqrt{\frac{\varepsilon_{33}+2B}{\varepsilon_{11}}}$$

becomes comparable to the thickness of the plasma layer, i.e., $\lambda_{pl}\approx a_{pl}$.

The results of the numerical study of the dispersion relation (6) for the case of a thick insulator layer ($k_1a_d\geqslant 1$) are shown in Fig. 1 and for a thin insulator layer ($k_1a_d\leqslant 1$) in Fig. 2. Increasing the thickness of the plasma layer makes the dispersion approach a form typical of a semi-infinite plasma–semi-infinite insulator model.[8] Note that in all the figures, the domain of existence of volume cyclotron waves lies above and to the right of the segments of the dispersion curves with reverse dispersion, denoted by the number 3. As the thickness of the plasma layer is reduced, the dispersion undergoes a transformation in the short wavelength region. Here the group velocity of the surface cyclotron waves, v_{gr}, in the segment with forward dispersion is considerably higher than in the previously studied metal–thin insulator layer–plasma structures.[19] For sufficiently small plasma regions, the group velocity of surface cyclotron waves approaches zero over the entire range of wavelengths and the

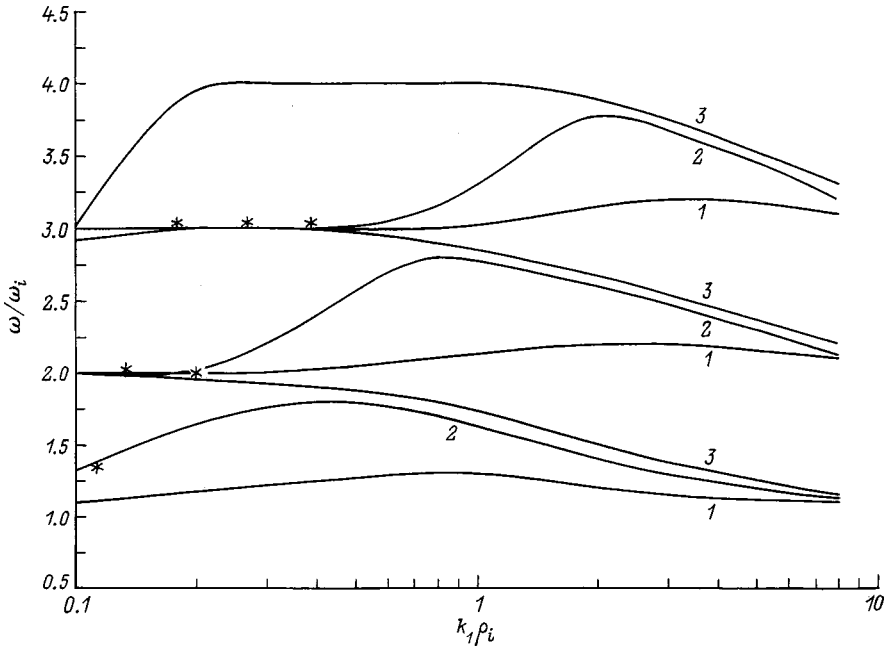

FIG. 1. The frequency of surface cyclotron waves normalized to the ion cyclotron frequency as a function of the wave vector normalized to the ion Larmor radius: $\omega_{pi}^2/\omega_i^2 = 2\times10^6$, $a_d/\rho_i = 10$; plasma layer thickness $a_{pl}/\rho_i = 10^{-6}$ (*1*), 10^{-5} (*2*), 10^{-4} (*3*).

resonance detuning h_i becomes smaller. This is an indication of a deterioration in the conditions for propagation of the waves being studied here.

A numerical study of the surface cyclotron wave dispersion relation (4) shows, increasing the thickness of the insulator layer causes the forward dispersion segment to shift toward longer wavelengths ($y_i \ll 1$). But it must be kept in mind that if the wavelength is long enough, the condition that they be slow, $k_3 v_{T\alpha} \ll (\omega - s|\omega_\alpha|)$, will be violated. As the thickness of the insulator layer is reduced, the group velocity of the wave increases in the forward dispersion region. In other words, there is an increase in the velocity at which the energy of the surface cyclotron wave is transported along the direction of propagation. As the dielectric permittivity of the insulator is increased, the dispersion properties of the surface cyclotron wave change in the same way as when the thickness of the insulator layer is increased.

We have also done a numerical study of the dependence of the dispersion of surface cyclotron waves on the magnitude of a constant external magnetic field. The results are shown in Fig. 3. For the parameters of experimental plasma and semiconductor electronics devices ($\omega_{pl}^2/\omega_i^2 \approx 2\times10^6$), the dispersion relation for surface cyclotron waves is similar the relation for these waves in a semi-infinite plasma–semi-infinite insulator (curve *3*). As ω_{pi}^2/ω_i^2 is reduced, the surface cyclotron wave dispersion relations squeeze toward the corresponding asymptotes $\omega = s\omega_i$. In controlled thermonuclear fusion experiments on stellarators, the characteristic plasma

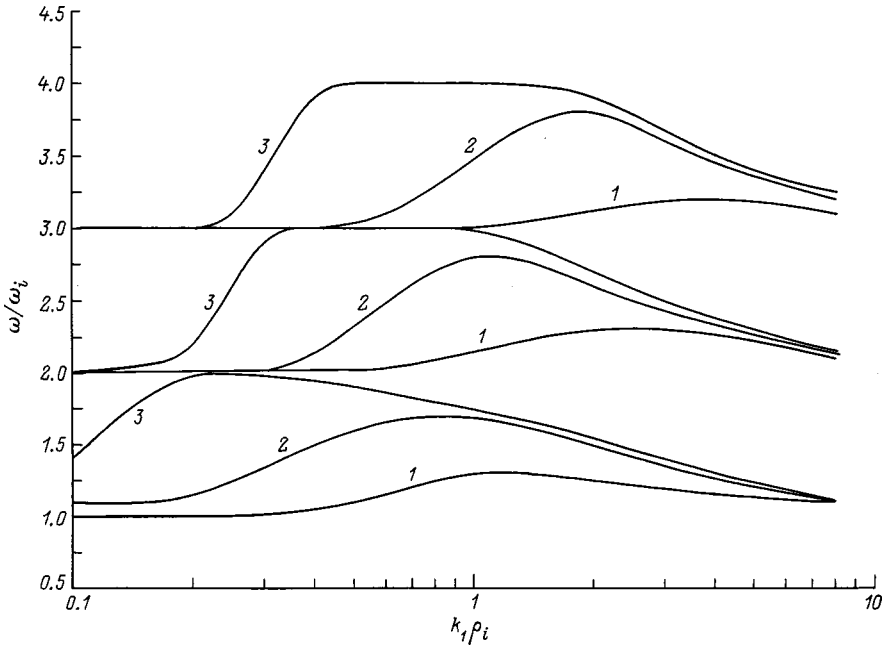

FIG. 2. As in Fig. 1, for $\omega_{pi}^2/\omega_i^2 = 2\times10^6$; $a_d/\rho_i = 10^{-3}$; $\alpha_{pl}/\rho_i = 10^{-3}$ (*1*), 10^{-2} (*2*), 10^{-1} (*3*).

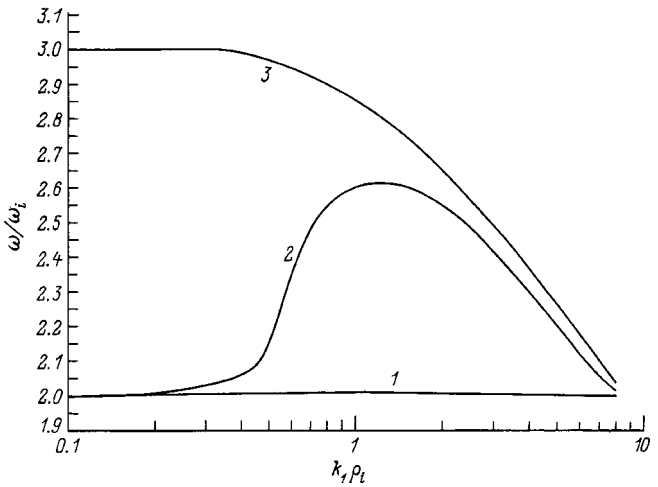

FIG. 3. As in Fig. 1 for $a_d/\rho_i=10$; $a_{pl}/\rho_i=10^{-3}$; $\omega_{pi}^2/\omega_i^2=50$ (*1*), 10^4 (*2*), 2×10^6 (*3*).

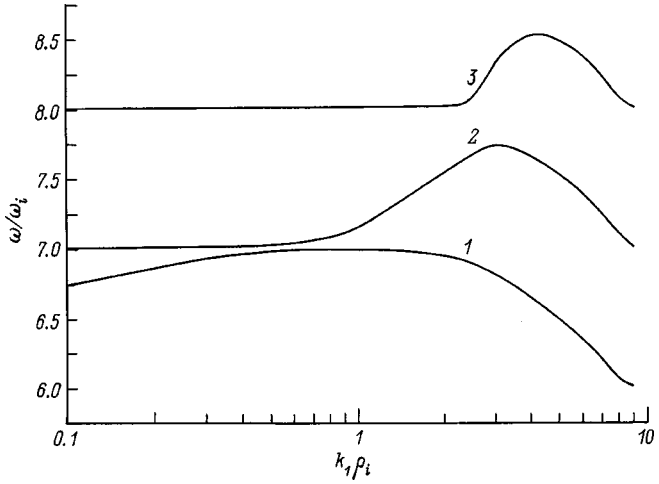

FIG. 4. As in Fig. 1. The conversion of ion surface cyclotron waves into a lower hybrid mode is shown for $\omega_{pi}^2/\omega_i^2=50$.

densities and magnetic fields in the divertor region vary over $n=10^{11}-10^{12}\,\mathrm{cm}^{-3}$ and $B_0=10^3-10^4\,\mathrm{G}$, with $\omega_{pi}^2/\omega_i^2 \approx 200$. Curve *1* was plotted for $\omega_{pi}^2/\omega_i^2=50$. This implies that for modelling the conditions of a divertor in thermonuclear experiments, the approximation of a semi-infinite plasma turn out to be incorrect and the approximation of a plasma layer with small transverse dimensions may have to be used. Here the depth of penetration of the surface cyclotron wave field into the plasma can exceed the dimensions of the plasma region.

A numerical calculation is in good agreement with the analytic solutions (7) and (8). In Fig. 1 the asterisks denote the surface cyclotron wave frequencies obtained from the analytic solutions of the dispersion relation (5) in the long wavelength limit. In the short wavelength limit the analytic solutions are valid for $y_i \geqslant (\omega_{pi}/\omega_i)^{4/3}$, so they do not show up on the scale of Fig.1.

A study of the dispersion relation (5) shows that it, as opposed to Eq. (4), has *N* solutions for a given harmonic number of the ion cyclotron frequency and a given wave number. But only one of these corresponds to a surface cyclotron wave propagating along the plasma–insulator interface, while the rest of the solutions appear because of our choice of a model for the inhomogeneity of the plasma density profile. These additional modes propagate along the boundaries of the plasma layers with different particle densities. The frequency of the surface cyclotron waves is lower for an inhomogeneous plasma than for a uniform plasma. The validity of this result of the numerical calculations for the surface cyclotron wave dispersion relation (5) is confirmed by the analytic solutions (7) and (8). Since the integrated plasma density is smaller for an inhomogeneous density profile than for a uniform plasma, according to Eqs. (7) and (8) the surface cyclotron wave frequency will also be lower.

Figure 4 shows some calculations based on Eq. (6) for an ion surface cyclotron wave in the case where it is converted into a lower hybrid mode subject to the condition $\omega^2(s=7)\approx\omega_e\omega_i$. Here it was assumed that $\omega_{pi}^2=50\cdot\omega_i^2$. A similar effect has also been observed with volume cyclotron waves.[3,4]

5. CONCLUSION

In this paper we have examined the dispersion properties of surface cyclotron waves at harmonics of the ion cyclotron frequency. These waves can exist only in dense plasmas with $\omega_{pi}^2>\omega_i^2$. This study has been done for the case of weak spatial dispersion of the medium along the normal to the interface between the media. Simple analytic expressions (7) and (8) for the eigenfrequencies of the ion surface cyclotron waves have been obtained for long ($y_i \ll 1$) and short ($y_i \gg 1$) wavelengths (compared to the ion Larmor radius) in the limiting cases of thick and thin insulator layers. As the thicknesses of the plasma and insulator layers are increased, the surface cyclotron wave dispersion curves develop a straight dispersion segment and the group velocity of the wave increases in this segment. In the limit of thin plasma ($|k_1|b \ll 1$) and thin insulator ($|k_1|a_d \ll 1$) layers, the ion surface cyclotron wave dispersion curves approach the asymptotic lines $\omega=s\omega_i$.

When the thicknesses of the plasma and insulator layers are very much greater than the penetration depth of the surface cyclotron wave field into the corresponding medium, the surface cyclotron wave dispersion curves resemble the dispersion curves for volume extraordinary cyclotron waves,[1–5] although their field structure and energy characteristics differ significantly. When the effect of the finite sizes of the plasma and insulator regions on the surface cyclotron wave dispersion relations are taken into account, a segment with a forward dispersion develops in which the group velocity of the wave can be controlled by changing either the thickness of the insulator layer or the size of the plasma region. Here the surface cyclotron wave frequency is lower than in the case of a semi-infinite plasma–semi-infinite insulator geometry. The penetration depth of the surface cyclotron wave into the plasma is very much greater than the wavelength. It should also be noted that a larger increase in the group velocity can be achieved by reducing the plasma region in the forward

dispersion segment than by reducing the thickness of the insulator layer. This implies that a metal–insulator–thin plasma layer will be unstable to changes in the external system parameters, e.g., parametric instabilities,[22] than a metal–thin insulator layer–plasma structure. Thus, when estimating the instability of a diverter plasma with respect to surface perturbations at cyclotron frequencies, the finite size of the plasma region should be kept in mind.

We believe that the surface cyclotron waves examined here can be used in various plasma technologies. Recently surface waves have been used for processing semiconductor surfaces. These technologies are being developed in Japan.[23,24] It has been found experimentally that in these processes, surface waves, as opposed to volume waves, play a controlling role. Traditionally, surface modes have been widely used for studying the electrical properties of solids.[7,9] Since the dispersion of surface cyclotron waves depends on the physical characteristics of insulator layers, surface cyclotron waves can be used for process control of insulator coatings on metal substrates. It appears to be possible to use surface cyclotron waves in electronics for creating rf generators and amplifiers. Here it would be possible to control the frequency of the wave perturbations by varying the charge carrier concentration.

[1] E. P. Gross, Phys. Rev. **82**, 232 (1951).
[2] A. G. Sitenko and K. N. Stepanov, Zh. Éksp. Teor. Fiz. **31**, 642 (1957) [Sov. Phys. JETP **4**, 512 (1957)].
[3] A. I. Akhiezer, I. A. Akhiezer, R. V. Polovin *et al.*, *Plasma Electrodynamics* [in Russian], Nauka, Moscow (1974), 720 pp.
[4] D. G. Lominadze, *Cyclotron Waves in Plasmas* [in Russian], Metsniereba, Tbilisi (1975), 224 pp.
[5] N. Krall and A. Trivelpiece, *Principles of Plasma Physics*, Mir, Moscow (1975), 528 pp.
[6] A. N. Kondratenko, *Plasma Waveguides* [in Russian], Atomizdat, Moscow (1976), 232 pp.
[7] V. M. Agranovich and D. L. Mill (Eds.), *Surface Polaritions*, North-Holland, Amsterdam (1982).
[8] A. N. Kondratenko, *Surface and Volume Waves in Finite Plasmas* [in Russian], Énergoatomizdat, Moscow (1985), 208 pp.
[9] S. Vucovich (Ed.), *Surface Waves in Plasma and Solids*, World Scientific, New York (1986).
[10] Ya. I. Kolesnichenko, V. V. Parail, and G. V. Pereverzev, in *Reviews of Plasma Physics*, Vol. 17, edited by V. V. Kadomtsev [Consultants Bureau, New York (1993) Énergoatomizdat, Moscow (1989), pp. 3–156].
[11] A. V. Longinov and K. N. Stepanov, in *Rf Plasma Heating* [in Russian], IPFAN SSSR, Gorky (1983), pp. 152–210.
[12] A. G. Efimov, A. G. Kirov, and V. P. Sidorov, *ibid.*, pp. 211–252.
[13] V. E. Golant and V. I. Fedorov, *RF Plasma Heating in Toroidal Fusion Devices*, Plenum, New York (1989), 194 pp.
[14] M. Moisan and J. Pelletin (Eds.), *Microwave Excited Plasmas*, Elsevier, Amsterdam (1992).
[15] *Microwave Discharges: Fundamentals and Applications*, NATO Advanced Study Institute, Vol. 302, Series B: Physics, Plenum, New York (1993).
[16] A. N. Kondratenko and V. M. Kuklin, *Principles of Plasma Electronics* [in Russian], Energoatomizdat, Moscow (1988).
[17] N. N. Beletskiĭ, A. A. Bulgakov, S. I. Khankina, and V. M. Yakovenko, *Plasma Instabilities and Nonlinear Effects in Semiconductors*, Naukova Dumka, Kiev (1984).
[18] N. A. Azarenkov, A. N. Kondratenko, and K. N. Ostrikov, Izv. Vuzov. Ser. Radiofizika **36**, 335 (1993).
[19] V. O. Girka, I. O. Girka, A. M. Kondratenko, and I. V. Pavlenko, Plasma Phys. **36**, 679 (1996).
[20] V. O. Girka, I. O. Girka, and I. V. Pavlenko, Plasma Phys. **58**, 31 (1997).
[21] M. Abramowitz and I. A. Stegun (Eds.), *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series, Vol. 55 (1964).
[22] V. O. Girka, V. I. Lapshin, and I. V. Pavlenko, *Contributed Papers of Intern. Symposium ''Plasma'97''*, Vol. 1, Opole (Poland) (1997), pp. 81–84.
[23] S. Nonaka, Jpn. J. Appl. Phys. **29**, 571 (1990).
[24] S. Nonaka, J. Phys. Soc. Jpn. **61**, 1449 (1992).

Translated by D. H. McNeill

Physical mechanisms for transition radiation of electromagnetic pulses

V. A. Balakirev and G. L. Sidel'nikov

Kharkov Physicotechnical Institute, National Science Center, 310108 Kharkov, Ukraine
(Submitted January 28, 1998; resubmitted June 3, 1998)
Zh. Tekh. Fiz. **69**, 90–95 (October 1999)

The excitation of electromagnetic pulses is examined in various generation schemes. The physical mechanism for excitation of the radiation is studied and the field from transition radiation by high current relativistic electron beams is analyzed. An analysis of the efficiency of the various generation schemes shows that the electron efficiency of beam radiators based on transition radiation may be quite substantial for beam parameters that are easily realized experimentally.
 [S1063-7842(99)01310-0]

INTRODUCTION

Research on the mechanisms for generation of high-power short electromagnetic pulses is of interest in a number of scientific and applied areas, in particular, for the development of a new generation of pulsed radar technology. Analytic studies and numerical simulations have shown[1,2] that the transition radiation owing to the interaction of a high-current burst of charged particles with a finite conducting body has a broad-band character. The output power and shape of the electromagnetic pulses are determined by the parameters of the current pulse and can be varied over wide limits.

High-power microwave generators employing high-current relativistic electron beams are intended to produce narrow band pulsed signals whose duration τ_p greatly exceeds the period T of the high frequency electromagnetic oscillations ($\tau_p/T \gg 1$) and are unsuitable for generating electromagnetic pulses with $\tau_p \leq T$. In fact, the generation of narrow band signals is the result of, and is maintained by, resonant energy transfer between the beam electrons and the excited electromagnetic field. The grouping of beam particles into bunches (modulation) and, therefore, the excitation of microwave radiation take place over times substantially longer than the period of the microwaves. This makes it clear that efficient generation of electromagnetic pulses by high-current beams requires the use of nonresonant (collisional) emission mechanisms, such as spontaneous coherent transition radiation, bremsstrahlung in an external magnetic field, or bremsstrahlung of short, pulsed beams in their own electromagnetic fields (pulsed radiation from a virtual cathode). The duration of the output electromagnetic pulses will be determined by the duration of the current in the burst and the output power, by the magnitude of that current and the energy of the burst. Short electron bursts with durations of $0.1-10$ ns, energies of $0.5-1$ MeV, and peak currents of $10-100$ kA can be obtained either by converting continuous electromagnetic beams into sequences of electron pulses (modulated beams)[3] or by direct generation of short and ultrashort high-current relativistic electron beams in high-current devices (exploding emission diodes).[4]

The fundamentally important feature of the transition radiation which develops during the motion of a charged particle in an electrically (magnetically) inhomogeneous medium is that the emission takes place with a constant particle velocity.[5] Because of this property, the transition radiation field from an electron beam is proportional to the beam current,[2] rather than to the time derivative of the current, as in traditional sources, such as an electric dipole, or to the second derivative with respect to time, as in the case of a magnetic dipole.

In this paper we present the results from a study of the generation of short duration, high-power electromagnetic pulses by transition radiation from high-current relativistic electron beams.

EXCITATION OF ELECTROMAGNETIC PULSES AS AN ELECTRON BURST INTERSECTS A CONDUCTING SCREEN

The model of the radiating device is illustrated in Fig. 1. An electron burst moves from a drift vacuum chamber *1* through a foil *2* into free space. The effect of the cylindrical boundary of the drift tube on the excitation of transition radiation is neglected. An electromagnetic pulse is formed as the electron burst intersects the conducting plane, which includes the foil *2* and a conducting screen *3*.

We shall assume that the radius of the screen is infinite. In addition, we neglect the change in the shape of the burst (the divergence of the beam particles) as it is transported into the interaction region. We write the current density of the pulse in the form

$$\mathbf{j} = -\mathbf{e}_z \frac{I_b}{2\pi} \psi(r) I\left(t - \frac{z}{V_0}\right),$$

where

$$I\left(t - \frac{z}{V_0}\right) = \frac{1}{I_b} \int_{-\infty}^{\infty} I_\omega e^{-i\omega(t - z/V_0)} d\omega,$$

$\psi(r)$ is a function describing the transverse beam profile, $I(t - z/V_0)$ specifies the shape of the bunch as a function of the longitudinal coordinate z and time t, V_0 is the velocity of the bunch, I_b is the peak beam current, and I_ω is the Fourier amplitude of the total current.

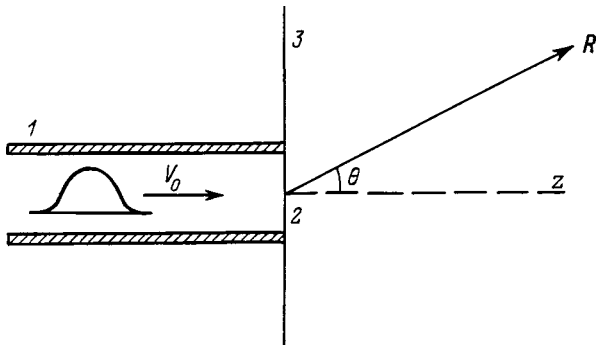

FIG. 1. A semi-infinite circular waveguide with an infinite flange.

The electromagnetic field excited by the bunch in the region $z>0$ (transition radiation) obeys Maxwell's equations, which are conveniently reduced to an inhomogeneous Helmholtz equation for the φ Fourier component of the magnetic field,

$$\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial H_\varphi}{\partial r}+\frac{\partial^2 H_\varphi}{\partial z^2}+\left(k_0^2-\frac{1}{r^2}\right)H_\varphi=\frac{4\pi}{c}\frac{\partial j}{\partial r}, \tag{1}$$

where $k_0=\omega/c$, ω is the angular frequency, and c is the speed of light.

For electron bunches with small transverse dimensions,

$$\frac{r_b}{c\tau_b\gamma}\ll 1,$$

where r_b is the radius of the bunch, τ_b is its duration, γ is the relativistic factor, and we can set $\psi(r)=\delta(r)/r$, where $\delta(r)$ is the delta function.

Equation (1) has cylindrical symmetry and is easily solved using the forward and inverse Fourier–Bessel transformations.

An asymptotic estimate of the field for $k_0R\gg 1$ yields an expression for the emitted pulse in the far field,[2]

$$H_\varphi=-\frac{2I_b}{cR}I(t-R/c)\frac{\beta\sin\Theta}{1-\beta^2\cos^2\Theta}, \tag{2}$$

where $\beta=V_0/c$, R is the distance from the coordinate origin to the observation point, Θ is the angle between the longitudinal axis of symmetry $(0z)$ and the vector $\mathbf{R}$.

Equation (2) yields the important result that the emitted pulse has exactly the same shape as the current pulse, i.e., the beam antenna for transition radiation has a broad-band character.

We now consider the efficiency with which the kinetic energy of the beam is converted into electromagnetic pulse energy. The total emitted power (power summed over all directions) is related to the parameters of the electron bunch by

$$P_{\rm rad}=2I^2(t-R/c)F(\beta)P_b\frac{I_b}{I_A}, \tag{3}$$

where $P_b=I_bmc^2(\gamma-1)/e$ is the peak power of the electron bunch, $I_A=mc^3/e=17\,\text{kA}$ is the Alfvén current, and

$$F(\beta)=\frac{1}{4(\gamma-1)}\left(\frac{1+\beta^2}{\beta}\ln\frac{1+\beta}{1-\beta}-2\right).$$

The total radiated energy is found by integrating Eq. (3) with respect to time. Thus,

$$W_{\rm rad}=\frac{2I_b^2}{c}(\gamma-1)F(\beta)\int_{-\infty}^{\infty}I^2(t)dt. \tag{4}$$

We define the electron efficiency of the beam radiator as the ratio of the emitted energy (4) to the kinetic energy of the bunch,

$$\eta=W_{\rm rad}/W_{\rm kin}, \tag{5}$$

where $W_{\rm kin}=\mathcal{Q}mc^2(\gamma-1)/e$ and $\mathcal{Q}=I_b\int_{-\infty}^{\infty}I(t)dt$ is the total charge in the bunch.

For a Gaussian profile

$$I(t)=\exp(-t^2/\tau_b^2), \tag{6}$$

integrating Eq. (4) gives $W_{\rm rad}=\sqrt{2}F(\beta)W_{\rm kin}(I_b/I_A)$. Thus, for the efficiency of a beam radiator, we have

$$\eta=\sqrt{2}F(\beta)I_b/I_A.$$

For a 1 MeV electron beam with a peak current of 8 kA, about 40% of the kinetic energy is converted into electromagnetic pulse energy.

EXCITATION OF ELECTROMAGNETIC PULSES DURING INJECTION OF AN ELECTRON BUNCH FROM A SEMI-INFINITE WAVEGUIDE

1. Basic equations. Let us consider a radiating system in the form of an ideally conducting semi-infinite cylindrical waveguide, whose end is covered by an ideally conducting diaphragm (foil) that is transparent to the beam particles. This device resembles that shown in Fig. 1 without the flange. The electron beam leaves the waveguide and enters free space through the diaphragm. The bunch has an annular cross section (tubular shape). We shall neglect the thickness of the bunch and approximate it by a surface current with density

$$\mathbf{j}=-\mathbf{e}_z\frac{I_b}{2\pi}\frac{\delta(r-r_b)}{r_b}I\left(t-\frac{z}{V_0}\right). \tag{7}$$

Equation (7) specifies a current pulse whose shape does not change as it propagates. The Fourier amplitude of the current density is given by

$$\mathbf{j}_\omega=-\mathbf{e}_z\frac{I_\omega}{2\pi}\frac{\delta(r-r_b)}{r_b}e^{i\omega z/V_0}. \tag{8}$$

The problem of finding the emitted electromagnetic pulse field reduces to solving the inhomogeneous system of Maxwell's equations for the Fourier amplitude of the electromagnetic field,

$$\operatorname{curl}\mathbf{E}_\omega-ik_0\mathbf{H}_\omega=0,\quad \operatorname{curl}\mathbf{H}_\omega-ik_0\mathbf{E}_\omega=\frac{4\pi}{c}\mathbf{j}_\omega.$$

We shall resolve the space into three regions: I — $\{r\leqslant r_b, z>0\}$, II — $\{a\geqslant r>r_b, z>0\}$, and III — $\{r>a, \infty>z>-\infty\}$, where a is the waveguide radius. The condi-

tions on the tangential components of the electromagnetic field at the boundaries between these regions and at the surface of the ideal conductor are

$$r=r_b, \quad E^{\mathrm{II}}_{z\omega}(z>0)=E^{\mathrm{II}}_{z\omega}(z>0),$$

$$H^{\mathrm{I}}_{\varphi\omega}(z>0)=H^{\mathrm{II}}_{\varphi\omega}(z>0)+\frac{2I_\omega}{cr_b}e^{i\omega z/V_0},$$

$$r=a, \quad E^{\mathrm{II}}_{z\omega}(z>0)=E^{\mathrm{III}}_{z\omega}(z>0),$$

$$H^{\mathrm{II}}_{\varphi\omega}(z>0)=H^{\mathrm{III}}_{\varphi\omega}(z>0),$$

$$E^{\mathrm{III}}_{z}(z<0)=0,$$

$$r\leqslant r_b, \quad E^{\mathrm{II}}_{r}(z=0)=0,$$

$$a\geqslant r>r_b, \quad E^{\mathrm{II}}_{r}(z=0)=0. \tag{9}$$

Because of the axial symmetry of the problem, we shall seek the radiated field in each region in the form of an integral superposition of cylindrical waves.

2. Mathematical model of the boundary value problem. Let us substitute the expression for the field components in the boundary conditions (9). According to the Wiener–Paley–Rapoport lemma,[6] the resulting system of integral equations determines the following coupled system of analytic functions in the complex k plane:

$$(e_1-e_2)J_0(\nu r_b)-e_3H_0^{(1)}(\nu r_b)=\varphi^+(k),$$

$$-i\frac{k_0}{\nu}[(e_1-e_2)J_1(\nu r_b)-e_3H_1^{(1)}(\nu r_b)]$$

$$=\frac{2I_\omega}{cr_b}\frac{1}{2\pi i}(k-h)^{-1}+\phi^+(k),$$

$$e_2J_0(\nu a)+(e_3-e_4)H_0^{(1)}(\nu a)=\Theta^+(k),$$

$$-i\frac{k_0}{\nu}[e_2J_1(\nu a)+(e_3-e_4)H_1^{(1)}(\nu r_b)]=\lambda^+(k),$$

$$e_4H_0^{(1)}(\nu a)=\kappa^-(k), \tag{10}$$

where $\nu=\sqrt{k_0^2-k^2}$.

In accordance with the radiation condition, we choose the branch of the double valued function $\nu(k)$, for which $\mathrm{Im}\,\nu\geqslant 0$ and $\mathrm{Re}\,\nu\geqslant 0$, where k is the longitudinal wave number, $J_n(\nu r)$ is the Bessel function, $H_n^{(1)}(\nu r)$ is the Hankel function, with $n=0,1$, and $e_\alpha(k)$ are the Fourier amplitudes of the fields, with $\alpha=1,2,3,4$. The functions φ^+, ϕ^+, Θ^+, and λ^+ are analytic in the upper half of the complex k plane and in the lower half of the complex κ^- plane, and $h=\omega/V_0+i0$. The boundary condition at the diaphragm $(z=0)$ implies

$$e_1(k)=e_1(-k), \quad e_2(k)=e_2(-k), \quad e_3(k)=e_3(-k). \tag{11}$$

Given Eq. (11), we can show (a consequence of the Wiener–Paley lemma) that there exists a coupled system, analogous to Eq. (10), with functions φ^-, ϕ^-, Θ^-, λ^- and κ^+ that are analytic, respectively, in the lower and upper halves of the k plane. Thus, the system of functional equations (10) subject to the conditions (11) is equivalent to the following system of Hilbert boundary value problems:

$$\varphi^+-\varphi^-=0, \tag{12}$$

$$\phi^+-\phi^-=-\frac{2I_\omega}{cr_b}\frac{k}{\pi i(k^2-h^2)}, \tag{13}$$

$$\kappa^+-\kappa^-=\Theta^+-\Theta^-, \tag{14}$$

$$\lambda^+-\lambda^-=-iZ_2(\Theta^+-\Theta^-), \tag{15}$$

$$Z_1(\Theta^++\Theta^-)+(Z_1-Z_2)(\kappa^++\kappa^-)$$

$$-i(\lambda^++\lambda^-)+\frac{r_b}{a}\frac{J_0(\nu r_b)}{J_0(\nu a)}[Z_1^b(\varphi^++\varphi^-)$$

$$-i(\phi^++\phi^-)]=\frac{2I_b}{ca}\frac{J_0(\nu r_b)}{J_0(\nu a)}\frac{h}{\pi(k^2-h^2)}, \tag{16}$$

where

$$Z_1=\frac{k_0}{\nu}\frac{J_1(\nu a)}{J_0(\nu a)}, \quad Z_2=\frac{k_0}{\nu}\frac{H_1^{(1)}(\nu a)}{H_0^{(1)}(\nu a)}, \quad Z_1^b=\frac{k_0}{\nu}\frac{J_1(\nu r_b)}{J_0(\nu r_b)}.$$

The solutions of Eqs. (12) and (13) are found trivially from the known jump at the contact surface (real axis). Taking Eq. (14) into account, as well, we obtain

$$\varphi^+\equiv\varphi^-\equiv 0, \quad \phi^+=-\frac{I_\omega}{cr_b}\frac{1}{\pi i(k+h)},$$

$$\phi^-=\frac{I_\omega}{cr_b}\frac{1}{\pi i(k-h)}, \quad \kappa^+=\Theta^+, \quad \kappa^-=\Theta^-. \tag{17}$$

The system of boundary value problems (15) and (16), with Eq. (17), can be reduced to the following singular integral equation for the function $\kappa\equiv\kappa^+-\kappa^-$:

$$2Z_1\hat{\mathcal{L}}\kappa-Z_2\hat{\mathcal{L}}\kappa-\hat{\mathcal{L}}Z_2\kappa=\frac{4I_\omega}{ca}\frac{J_0(\nu r_b)}{J_0(\nu a)}\frac{h}{\pi(k^2-h^2)}. \tag{18}$$

Here $\hat{\mathcal{L}}$ is the Cauchy operator. It is not easy to find the general solution to the integral Eq. (18). However, in the limiting case of a small waveguide radius, i.e.,

$$k_0a\ll 1 \tag{19}$$

Eq. (18) can be greatly simplified and solved approximately. In fact, neglecting the term proportional to $Z_1\approx k_0a$ in Eq. (18), we shall have

$$Z_2\hat{\mathcal{L}}\kappa+\hat{\mathcal{L}}Z_2\kappa=-\frac{4I_\omega}{ca}\frac{h}{\pi(k^2-h^2)}. \tag{20}$$

Using the notation $U=Z_2\kappa$ and the Sohocky formula, we reduce Eq. (20) to an inhomogeneous Hilbert boundary value problem for the half plane with contour $\{k\colon \mathrm{Im}\,k=0\}$,

$$U^+=-Z_2\kappa^--\frac{2I_\omega}{ca}\frac{h}{\pi(k^2-h^2)}, \tag{21}$$

where U^+ and κ^- are the unknown functions.

The solution of Eq. (21) for κ^- has the form

$$\kappa^- = \frac{2I_\omega}{cak_0}\frac{X^-(k)}{X^+(h)}\frac{\sqrt{(k_0-k)(k_0+h)}}{2\pi i(k-h)}.$$

In the limiting case (19), the expressions for the functions $X^\pm(k)$ have the form[7]

$$X^+ = \frac{\sqrt{\ln(i/\nu k_0 a)}}{\sqrt{k_0+k}\,\ln[2i/\nu a\sqrt{2k_0(k_0+k)}]},$$

$$X^- = \frac{\sqrt{k_0-k}\,\ln[2i/\nu a\sqrt{2k_0(k_0-k)}]}{i\sqrt{\ln(i/\nu k_0 a)}},$$

$$\nu = 1.781. \tag{22}$$

The magnetic field for $r>a$ is given by

$$H^{\mathrm{III}}_{\varphi\omega} = -\frac{2I_\omega}{ca}\int_\infty^\infty \frac{X^-(k)}{X^+(h)}\frac{H_1^{(1)}(\nu r)}{H_0^{(1)}(\nu a)} \times\sqrt{\frac{k_0+h}{k_1+k}}\frac{e^{ikz}}{2\pi(k-h)}dk. \tag{23}$$

Restricting ourselves to the principal term in the expansion of the field for $k_0R\gg 1$ and using the representation (22) for the functions $X^\pm(k)$, we shall have

$$H^{\mathrm{III}}_{\varphi\omega} = \frac{I_\omega}{c}\frac{e^{ik_0R}}{R}\Lambda(\Theta,\omega)\cdot Y(\Theta), \tag{24}$$

where

$$Y(\Theta) = \frac{\beta\sin\Theta}{1-\beta\cos\Theta} + \frac{\sin\Theta}{1+\cos\Theta}$$

and

$$\Lambda(\Theta,\omega) = \ln\frac{2i}{\nu k_0 a\sqrt{2(\beta^{-1}+1)}} \Big/ \ln\frac{2i}{\nu k_0 a\sqrt{2(1+\cos\Theta)}}.$$

The first term originates in the emission by the charged particle beam, while the second corresponds to the emission produced by the current induced by the beam in the surface of the conductor.

The electron beam emits in the forward direction at an angle of $\Theta\approx 1/\gamma$ to the axis. The induced current radiates backwards. The maximum emission from the induced current is at an angle of $\theta=\pi$. This is because of the directional properties of a thin cylindrical conductor. A gaussian current pulse (6) with a duration $\tau_b=1$ ns was chosen for the specific calculations.

A numerical analysis of the electromagnetic pulses showed that the duration of the emitted pulse increases slightly compared to the current pulse for angles Θ close to π. At the same time, because of the weak (logarithmic) dependence of the function $\Lambda(\theta)$ for angles Θ not too close to π, the shape of the emitted electromagnetic pulse is extremely close that of the applied current pulse and essentially reproduces it. The total energy emitted in the electromagnetic pulse can be calculated using the formula

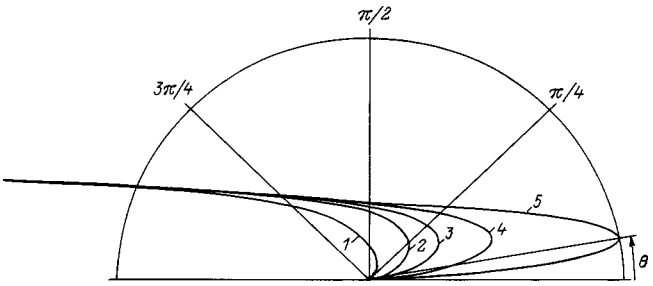

FIG. 2. Directional diagram of the energy emitted in electromagnetic pulses for $\gamma=1.05$ (*1*), 2 (*2*), 3 (*3*), 5 (*4*), and 10 (*5*).

$$W_{\mathrm{rad}} = \frac{I_b^2}{c}\frac{\tau_b}{2\pi}\int_0^\pi D(\Theta)d\Theta, \tag{25}$$

where

$$D(\Theta) = \sin\Theta\cdot Y^2(\Theta) \times\int_{-\infty}^{\infty} d\tau\left[\mathrm{Re}\int_0^\infty e^{-\Omega^2/4-i\Omega\tau}\Lambda(\Theta,\Omega)d\Omega\right]^2,$$

$$\Omega\equiv\omega\tau_b, \quad \tau\equiv\frac{t}{\tau_b}.$$

Figure 2 shows the energy of the radiation pulse as a function of the observation angle for different values of γ. The directional diagram has a clear maximum at small angles for relativistic energies of the driver electron bunch. As Θ increases, for angles greater than $\pi/2$ the angular distribution of the energy emitted by the electromagnetic pulse is essentially independent of γ. This is because the backward emission is formed on the surface of the thin cylindrical conductor by the current of induced free charges moving in a direction opposite the bunch. In terms of the physical model considered here, the directional diagram of the energy emitted in the electromagnetic pulse has a singularity at observation angles close to π, but the total emitted energy (the energy integrated over all angles) remains finite, since this singularity is integrable. The curves in Fig. 2 have been plotted on a scale corresponding to the peak output energy for $\gamma=10$.

The emitted power corresponding to the peak pulse as a function of polar angle has the same appearance as the corresponding plot of the emitted energy in Fig. 2.

We now determine the efficiency of the proposed scheme for a beam driven electromagnetic pulse source. According to Eq. (5), for a gaussian bunch (6), we have

$$\eta = \frac{1}{2\pi\sqrt{\pi}}\frac{I_b}{I_A}\frac{\int_0^\pi D(\Theta)d\Theta}{\gamma-1}.$$

The radiator efficiency as a function of the energy in the exciting bunch for a peak current $I_b=3$ kA is plotted in Fig. 3. The ratio (ξ) of the energy of the electromagnetic pulse emitted in the forward direction to the energy emitted from

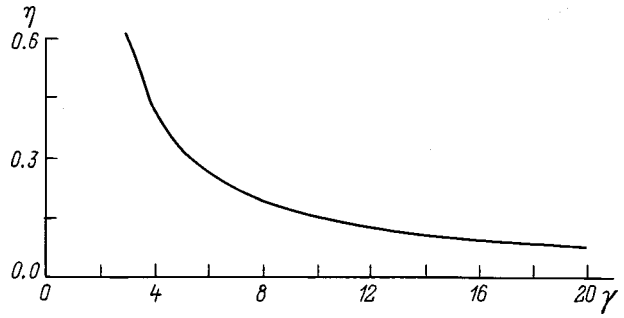

FIG. 3. The radiator efficiency as a function of the energy in a bunch.

the surface of the conductor by the induced current is plotted in Fig. 4 as a function of the relativistic factor γ (the kinetic energy of the bunch).

If the forward emission, into $\Theta<\pi/2$, is of interest then in Eq. (24) one can set $\Lambda(\Theta)=1$ for a relativistic beam. Integrating by parts, we obtain a simple expression for the output pulse,

$$H_{\varphi}^{\mathrm{III}}=\frac{I(t-R/c)}{cR}Y(\Theta). \tag{26}$$

As in the case of a conducting half space, the forward emitted pulse has the same shape as the beam current pulse.

EXCITATION OF ELECTROMAGNETIC PULSES IN A COAXIAL LINE

A method has been proposed[8,9] for exciting short (picosecond) electromagnetic pulses with powers $\gtrsim 10^9$ W by scanning a relativistic electron beam on the inner surface of a coaxial conductor without rigid requirements on the synchronism between the scan rate and the velocity of the charge wave. We shall show that electromagnetic pulses can be excited extremely efficiently based on the transition radiation effect in a special device, a coaxial converter.

Figure 5 is a sketch of a possible device of this type. A pulsed electron beam from a drift tube in the form of a circular waveguide is incident on a center conductor that is coaxial with the waveguide. As a result of the transition radiation from the beam, a *TEM* field pulse is excited in the coaxial line.

We shall consider an electron bunch with an annular cross section (7). The radius of the bunch is smaller than that

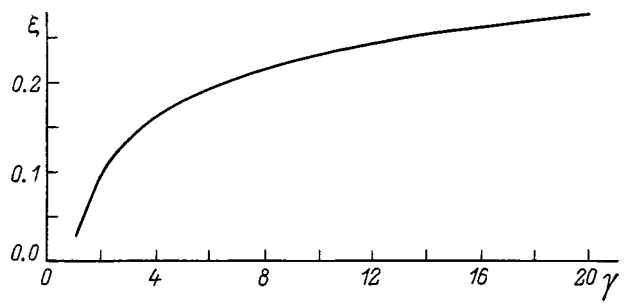

FIG. 4. The ratio of the energy emitted in the forward direction ($\Theta<\pi/2$) to the energy emitted by the induced current ($\Theta>\pi/2$) as a function of the parameter γ.

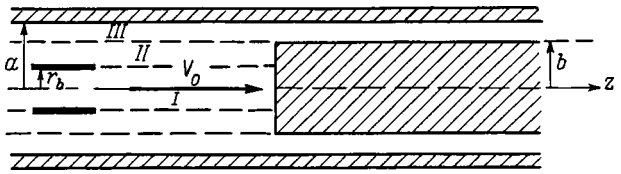

FIG. 5. Coaxial line with a semi-infinite cylindrical rod.

of the inner conductor. We divide the free space inside the waveguide into three regions. To determine the radiation pulse, we shall use the Hilbert boundary value problem developed in the previous section. We seek the field in each of the regions I — $\{r<r_b,\ z<0\}$, II — $\{b>r>r_b,\ z<0\}$, and III — $\{a>r>b,\ \infty>z>-\infty\}$ in the form of integrals with respect to the longitudinal wave numbers of cylindrical waves. Here a is the waveguide radius and b is the radius of the central conductor. Using the continuity conditions for the tangential components of the field at the interfaces of regions I, II, and III, we obtain the following system of functional equations

$$\begin{aligned}
&(e_1-e_2)J_o(vr_b)-e_3\Delta_0(vr_b,vb)\\
&\quad=\varphi^-+\frac{ik_0}{v}[(e_1-e_2)J_1(vr_b)-e_3\Delta_1(vr_b,vb)]\\
&\quad=-\frac{2}{c}\frac{I_\omega}{r_b}\frac{1}{2\pi i(k-h)}+\phi^-,\\
&e_2J_0(va)+(e_3-e_4)\Delta_0(va,vb)=\Theta^-,\\
&-i\frac{k_0}{v}[e_2J_1(va)+(e_3-e_4)\Delta_1(va,vb)]=\lambda^-,\\
&e_4\Delta_1(va,vb)=\kappa^+.
\end{aligned} \tag{27}$$

Then, proceeding as in the previous section, we reduce the system of functional equations to a singular integral equation for the function κ and find a solution in the quasistatic approximation (19). This case is of greatest interest, since the contribution of the waveguide modes to the radiation field is negligible. The bulk of the beam energy goes into exciting a *TEM* field in the coaxial waveguide.

The solution of the integral equation makes it possible to determine the Fourier amplitude $e_4(k)$ and, ultimately, to find the electromagnetic field in the coaxial line,

$$\begin{aligned}
H_{\varphi\omega}^{\mathrm{III}}=&-i\frac{I_\omega}{ca}\int_{-\infty}^{\infty}\frac{X^{-1}(h)}{X^+(k)}\frac{\Delta_1(va,vb)}{\Delta_0(va,vb)}\\
&\times\sqrt{\frac{k_0-h}{k_0-k}}\frac{e^{ikz}}{2\pi(k-h)}dk.
\end{aligned} \tag{28}$$

The functions $X^{\pm}(k)$ (the solutions of the homogeneous problem) can be approximated by

$$\begin{aligned}
X^+&=\sqrt{\frac{i}{\ln(b/a)}}\frac{1}{\sqrt{(k_0+k)a}},\\
X^-&=\sqrt{\frac{\ln(b/a)}{i}}\sqrt{(k_0-k)a}.
\end{aligned} \tag{29}$$

Leaving out the calculations, here we give an expression for the excited TEM pulse,

$$H_{\varphi}^{\mathrm{III}}=\frac{2I(t-z/c)}{cR}. \tag{30}$$

This formula implies that the shape and amplitude of the TEM pulse are determined only by the current in the bunch. The power transported in the TEM pulse is given by

$$P_{\mathrm{rad}}=\frac{2I^2}{c}\ln(a/b). \tag{31}$$

The total radiated energy is found by integrating Eq. (31) with respect to time. According to Eq. (5), for a Gaussian beam (6), we have

$$\eta=\frac{\sqrt{2}\ln(a/b)}{\gamma-1}\frac{I_b}{I_A}.$$

In particular, for $a\approx 2.7b$, a beam particle energy of 1 MeV ($\gamma=3$), and a peak current of 10 kA, the beam energy is converted into TEM-pulse energy with an efficiency of 40%.

CONCLUSION

In this paper we have proposed some nontraditional approaches to creating high-power electromagnetic pulse sources in the gigawatt range based on the coherent transition radiation of a bunch or sequence of bunches of electrons. Various schemes for exciting electromagnetic pulses have been examined. These studies have shown that transition ration is an efficient mechanism for generating electromagnetic pulses. The shape of the output pulse and its duration are close to those of the electron beam current pulse. The efficiency of radiators using beams with currents of 5–20 kA and energies of 0.5–1 MeV can be 30–60%. Magnetobremsstrahlung radiation from short pulse high-current relativistic electron beams in an external magnetic field and the radiation generated by a short pulse high-current relativistic electron beam during formation of a virtual cathode are also promising mechanisms for generating electromagnetic pulses.

[1] V. A. Balakirev and G. L. Sidel'nikov, *Transition Radiation of Modulated Electron Beams in Inhomogeneous Plasmas* [in Russian], KhFTI, Kharkov (1994), 104 pp.
[2] V. A. Balakirev and G. L. Sidel'nikov, Pis'ma Zh. Tekh. Fiz. **22**(10), 45 (1996) [Tech. Phys. Lett. **22**(5), 407 (1996)].
[3] M. Friedman, V. Serlin, Y. Y. Lau, and J. Krall, in *Proceedings of the 8th International Conference on High-Power Particle Beams (BEAMS'90)*, Novosibirsk (1990), Vol. 1, p. 53.
[4] V. E. Chaĭka and I. M. Goncharuk, *Materials from the Conference on Microwave Technology and Satellite Communications* [in Russian], Sevastopol (1994), Vol. 2, p. 393.
[5] V. L. Ginzburg and V. N. Tsytovich, *Transition Radiation and Transition Scattering* [in Russian], Nauka, Moscow (1984), 358 pp.
[6] I. M. Rapoport, DAN SSSR **59**, 1403 (1948).
[7] L. A. Vaĭnshteĭn, *Diffraction Theory and the Method of Factorization* [in Russian], Sov. Radio, Moscow (1966), 431 pp.
[8] L. N. Kazanskiĭ and A. A. Rukhadze, Pis'ma Zh. Tekh. Fiz. **20**(3), 26 (1994) [Tech. Phys. Lett. **20**(2), 101 (1994)].
[9] L. N. Kazanskiĭ and A. A. Rukhadze, Zh. Tekh. Fiz. **66**(3), 107 (1996) [Tech. Phys. **41**, 289 (1996)].

Translated by D. H. McNeill

Spatial harmonics of the field in a quadrupole mass filter with circular electrodes

D. J. Douglas

University of British Columbia, 2036 Main Mall, Vancouver, B. C. V6T 1Z1, Canada

T. A. Glebova, N. V. Konenkov, and M. Yu. Sudakov

S. A. Esenin Ryazan State Pedagogical University, 390000 Ryazan, Russia

(Submitted May 19, 1998)
Zh. Tekh. Fiz. **69**, 96–101 (October 1999)

The amplitudes of the spatial harmonics of the field are calculated for given displacements of the electrodes in a quadrupole mass filter relative to their optimum positions. The cases of radial and angular displacements of the electrodes, asymmetries in voltage supplied to the electrodes, and nonidentical rod diameters are considered. These data are required for calculating the actual shape of the mass peaks and will be useful in designing mass filters.
 [S1063-7842(99)01410-5]

INTRODUCTION

Quadrupole mass spectrometers are widely used in various techniques for analyzing the composition and structure of matter because of their great advantages in terms of low cost and characteristics: mass range, resolution, and sensitivity. The design flexibility of the quadrupole analyzer means that it can be used effectively in tandem mass spectrometry (MS–MS), mass spectrometry of inductively coupled plasmas (ICP MS), and secondary ion mass spectrometry (SIMS), as well as for selective detection in gas and liquid chromatography.[1,2]

The main requirement for a quadrupole mass spectrometer is that the mass peak should have high quality (without regular dips in the transmission profile), when it is necessary to separate ions of a single mass for a long time: parent ions in MS–MS, layer-by-layer elemental analysis in SIMS, and isotopic analysis in ICP MS.

The splitting of a mass peak is explained by resonant enhancement of the amplitude of the oscillations of the ions in harmonics of the field, when part of the ions drift to a rod during scanning and reach the detector.[3–7] The conditions for observing nonlinear resonances were first obtained by Lammert.[2] A quantitative relationship has been obtained between the limiting resolution and the permissible electrode loss for the first stability region.[3] A general theory of nonlinear resonances in weakly nonlinear quadrupole fields has been developed.[4,5] The conditions for observing nonlinear resonances as applied to quadrangular stability $(a, q \approx 3)$ have been obtained in Ref. 7, where it has been shown that resonant oscillations are not excited at the lowest harmonics of the field for high resolutions. This makes it possible to attain resolutions of more than 9000 and 5000, respectively, in simple electrode structures with circular rods operating in the second $(a \approx 0,\ q \approx 7.55)$[8] and third $(a, q \approx 3)$[9] stability regions.

It has been shown[4] that detailed knowledge of the harmonics of the analyzer field is required for the theoretical description of the shape of the mass peak. The deviation of the electric field in a quadrupole mass filter from hyperbolic is caused by errors in fabricating and assembling the electrodes.

In this paper we calculate the spectral composition of the harmonics of the field of a quadrupole mass filter as a function of the displacement of circular electrodes from their optimum positions. Knowledge of these dependences makes it possible to compute the effect of the analyzer geometry on the shape of a mass peak and, thereby, to determine the tolerances for the fabrication and assembly of the electrodes when they are used in the various stability regions of a quadrupole mass filter.

STATEMENT OF THE PROBLEM

For the case of a two-dimensional field, when the rods (electrodes) of the analyzer are parallel (Fig. 1), the spectral composition of the spatial harmonics of the field $\varphi(r, \theta)$ can be described using the series

$$\varphi(r,\Theta) = \sum A_N (r/r_0)^N \times \cos[N(\Theta - \Theta_0)](U + V\cos\omega t), \qquad (1)$$

where r and Θ are the polar coordinates, A_N are the harmonic amplitudes, Θ_0 is the initial phase, N is the harmonic number, $\pm(U + V\cos\omega t)$ are the supply voltages applied to opposite pairs of electrodes, and r_0 is the ''field radius''[4] (the radius of the circle inscribed between the electrode tips).

A linear working field corresponds to harmonic $N = 2$, for which the field components $E_x \propto x$ and $E_y \propto y$. When harmonics (1) are present, the field components are nonlinear and coupled. For our case of a two-dimensional field (1), we restrict ourselves to field distortions caused solely by parallel displacements of the electrodes. Noncylindrical, barrel-shaped, roughness, and nonparallel distortions of the electrodes[10] are neglected in this approach.

Our task is to calculate the electric field created by four nonidentical electrodes positioned in an external metal shell of radius R at zero potential, as shown in Fig. 1. The multi-

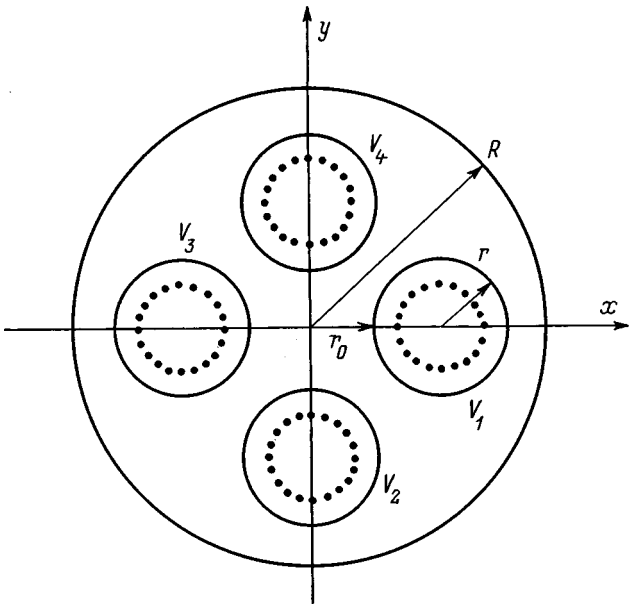

FIG. 1. Structure of the electrodes in a quadrupole mass filter and the location of the elementary charges (denoted by dots in circles).

poles are calculated in order to study the harmonic composition of the field, i.e., the amplitudes $A_N(\Delta r)$ for specified displacements Δr of the electrodes by amounts $\pm\Delta r=(0-10^{-2})r_0$.

METHOD FOR CALCULATING THE ELECTRIC FIELDS

We shall consider a system consisting of four cylindrical electrodes of infinite length positioned parallel to one another. With these restrictions, the electric field is independent of the longitudinal coordinate and the problem reduces to solving the Laplace equation in two dimensions. Despite this significant simplification, the problem still cannot be solved analytically in general. Thus, to calculate the fields in this paper we shall use a numerical method, the ''method of equivalent charges.''[11]

The essence of this method is the following. The object is to find a function satisfying the Laplace equation in some region G bounded by a finite number of closed surfaces on which it takes fixed values (Dirichlet problem). We shall use a function which is known to satisfy the Laplace equation, the potential of a point charge. We place K elementary charges outside the boundaries of region G at fixed points, whose coordinates are known. We take the same number K of charges on the boundaries of region G and require that the resulting potential of all K charges at these points satisfy the boundary conditions. This gives us a system of K linear equations for determining the magnitudes of the charges. On determining the latter, we obtain a potential function which is a superposition of the potentials of the elementary charges and clearly satisfies the Laplace equation and has no singularities in the region G, since the elementary charges lie outside the region and it also satisfies the boundary conditions at a finite number K of points on the boundary of the region. The accuracy with which this function approaches the true solution can be estimated by calculating its values at other points on the boundary of the region.

For the two-dimensional problem, a point charge is an infinite charged filament. The potential function of a filament in the perpendicular plane is equal to the real part of the analytic complex function $u_1(z)=\lambda_1\ln(z-z_0)$, where λ is the linear charge density of the filament, $z=x+iy$ are the cartesian coordinates of a point, and $z_0=x_0+iy_0$ is the location of the charge. Here and in the following we represent the two-dimensional potential by complex numbers, since, according to the Cauchy–Riemann conditions, the real and imaginary parts of any analytic function satisfy the Laplace equation (i.e., the function is harmonic). It is known that the potential of a point charge located inside a grounded circle is the superposition of the potentials of two point charges, one of which is the original charge at the point $z_0=x_0+iy_0$ while the other is its image with respect to the grounded circle lying on the same ray at the point $z_1=R^2/z_0^*$. In order to automatically account for the grounded metal shell of radius R with its center at the coordinate origin in our problem, it is necessary to take a potential of the ''elementary charge'' in the form of a mirror pair as the real part of the function

$$u_2(z,z_0)=\lambda[\ln(z-z_0)-\ln(z-R_c^2/z_0^*)+\ln(R_c/z_0^*)]. \tag{2}$$

The last term in this formula ensures that the total potential of the pair equals zero on the circle $z=R\cdot e^{i\varphi}$. In the case of a mass filter with ideal cylindrical symmetry of the electrode placement and absolutely symmetric power supplies, the electric field has an additional mirror symmetry with respect to the bisector of the coordinate angles. In this case, it is possible to limit the field calculation to just the first quadrant of the plane and the potential of the ''elementary charge'' has to be taken in the form of a superposition of fields (2) which are mirror symmetric with respect to one another at the coordinate axes,

$$u_4=u_2(x,y;x_0,y_y)+u_2(x,y;-x_0,-y_0) -u_2(x,y;y_0,x_0)-u_2(x,y;-y_0,-x_0). \tag{3}$$

The potential (3) goes to zero on the coordinate axes, as well as on the circle R.

In the case of a mass filter with symmetrically positioned electrodes, in order to determine the field by the method of equivalent charges we placed a finite number $K=60-120$ of elementary charges of type (3) inside a cylinder r at distances of $(0.5-0.8)\cdot r$ from its center. For arbitrary position and supply of the electrodes, charges of type (2) lie within each of the four electrodes (Fig. 1).

Determining the amplitude of the spatial harmonics of the electric field A_N using the complex representation of the potentials reduces to expanding the potential in the Taylor series

$$U(x,y)=\mathrm{Re}\left[\sum_{N=0}^{\infty}A_N\cdot(z/r_0)^N\right] =\mathrm{Re}\left[\sum_{N=0}^{\infty}A_N\cdot(r\cdot e^{i\varphi}/r_0)^N\right]. \tag{4}$$

The magnitudes of the charges are determined by a system of linear equations which is obtained by satisfying the boundary conditions at a finite number of points. Having determined the magnitudes of the charges and using the known position of the logarithmic function in the Taylor series (2), we find the amplitudes A_N of the spatial harmonics of the resulting field in the form of a superposition of the fields created by the individual charges.

RESULTS AND DISCUSSION

1) *Optimum electrode positions.* In this problem the calculation is done for a system with symmetric position and power supply of the electrodes, $V_4=V_2=V$ and $V_1=V_3=-V$. Because of the symmetry of the boundary conditions, the amplitudes of the odd order spatial harmonics are zero. In the working region of the mass filter, the field is given by

$$\varphi(x,y)=V\varphi_2(x,y)=V\cdot(A_2r^2\cos(2\varphi)+A_6r^6\cos(6\varphi)+A_{10}r^{10}\cos(10\varphi)+\ldots). \quad (5)$$

In this case the potential contains the spatial harmonics $N=6, 10, 14, 18\ldots$, in addition to the fundamental $N=2$. The optimum position for circular electrodes with radius r is determined by the structural parameter r_0 for which the amplitude of the sixth harmonic $A_6=0$[12]. r_0 also depends on the radius R of the cylindrical metal shell. The latest studies[12] show that there an optimum value of $r/r_0=1.14511$, which depends weakly on the shell radius R.

In the calculations we chose the practical case of a shell radius $R=4r_0$. The criterion for the required accuracy of the field distribution calculation was the magnitude of the deviation of the potential at the electrode surfaces. The specified error $\Delta\varphi/\varphi\leqslant10^{-8}$ in the calculations was achieved by choosing the positions for the charges indicated by the dots in Fig. 1 within the interval $(0.5-0.8)r$ and with $K\geqslant50$ charges. All of the following calculations were done for $K=60$ with charges located on a circle of radius $0.6r$, which required solution of a system of $4k$ linear equations.

Our method also yielded the ''magic'' value[12,13] of $r/r_0=1.145111$ for $R=4r_0$; this value had been found earlier by conformal mapping. In the absence of a grounded cylinder ($R\to\infty$). this ratio equals 1.1451081. We then examined the effect of symmetric drift of the four rods from their optimum positions on the amplitudes A_N of the field harmonics.

Figure 2 shows plots of the coefficients A_N for $N=2, 6, 10, 14, 18$ as functions of the displacement Δr. Here and in the following, the unit of length is taken to be r_0, i.e., the displacement Δr is represented by $\Delta r/r_0$. We limit ourselves to the 18th harmonic, since resonances in the higher stability zones of a quadrupole mass filter can be observed there.[7] We note as an example that $\Delta r=5\cdot10^{-3}$ corresponds to 1 μm for a standard field radius $r_0=5$ mm. Figure 2 implies that a symmetric, simultaneous shift of the four rods over fairly wide limits of $1.12<\Delta r/r_0<1.18$ has little effect on the amplitudes of the harmonics of the field with $N\geqslant10$. Thus, one of the major tasks in the technology of fabricating quadrupole analyzers is to ensure maximal symmetry in the position of the four electrodes. In this case, resonant oscillations of the ions can show up only in the harmonics with $N=10, 14, 18$ for resolutions >100, and the depth of the dips in the mass peaks will depend weakly on the position of the rods for a quadrupole mass filter operating in high stability zones.[7]

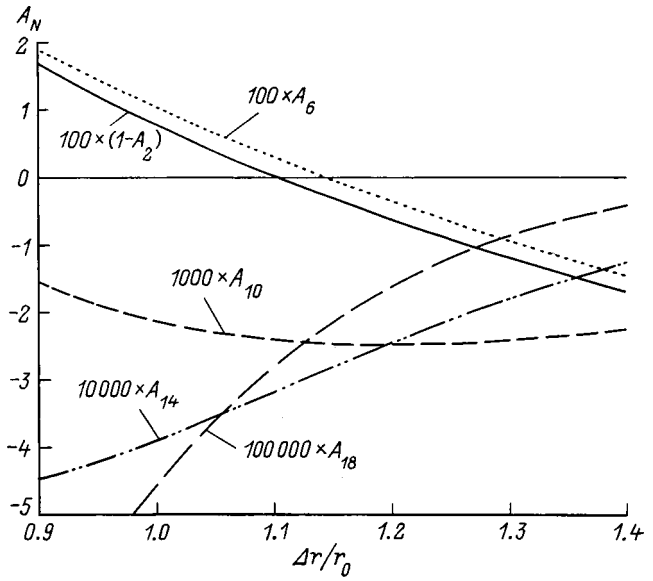

FIG. 2. The amplitudes A_N of the spatial harmonics as functions of the displacement Δr for a symmetric shift of the four cylindrical electrodes.

2) *Asymmetry in power supply to the electrodes.* Experiments[13] show that an asymmetry in the power supply to opposite electrodes leads to an asymmetry in the shape of a mass peak and a reduction in the throughput of a quadrupole mass filter. A sharp drop in the output signal is observed with an imbalance of 2–4% in the supply voltages. Thus, it is important to know the spectral composition of the harmonics owing to asymmetric supply. The method we have proposed for calculating the electric field makes it possible to analyze the case of asymmetric power supply in a mass filter for arbitrary electrode potentials V_1, V_2, V_3, and V_4 (Fig. 1), but, in practice, equal potentials are applied to the electrode pairs 1, 3 and 2, 4. Thus, the situation where $V_1=V_3=V$ and $V_2=V_4=-V-\Delta V$ is of practical interest. In this case, based on the principle of superposition we obtain

$$\varphi(x,y)=-0.5\Delta V\varphi_0(x,y)+(V+0.5\Delta V)\varphi_2(x,y), \quad (6)$$

where $\varphi_0(x,y)$ is the field produced in the system with identical potentials on all four electrodes, i.e., $V_1=V_2=V_3=V_4=1$, while the field $\varphi_2(x,y)$ occurs for quadrupole feed with $V_1=V_3=1$ and $V_2=V_4=-1$. The spectral composition of the field $\varphi_2(x,y)$ was studied in the previous section. According to Eq. (6), an asymmetry in the power supply supplements this field with a component $\varphi_0(x,y)$, which does not change upon rotation by 90° and has an amplitude that is directly proportional to the magnitude of the asymmetry, $-0.5\Delta V$. This result is exact, since it follows only from the superposition principle. Thus, for a complete analysis of the effect of the asymmetry, we study the spectral composition of the field $\varphi_0(x,y)$ as a function of r/r_0. As opposed to the field $\varphi_2(x,y)$, the symmetric component $\varphi_0(x,y)$ depends strongly on the radius R of the grounded electrode (in

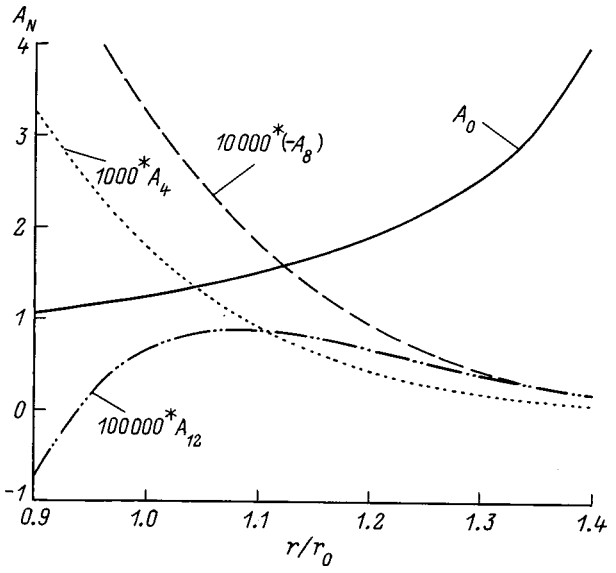

FIG. 3. The multipole components of the field induced by an asymmetry in the supply voltages to the electrodes as a function of the relative shift r/r_0 of the four electrodes.

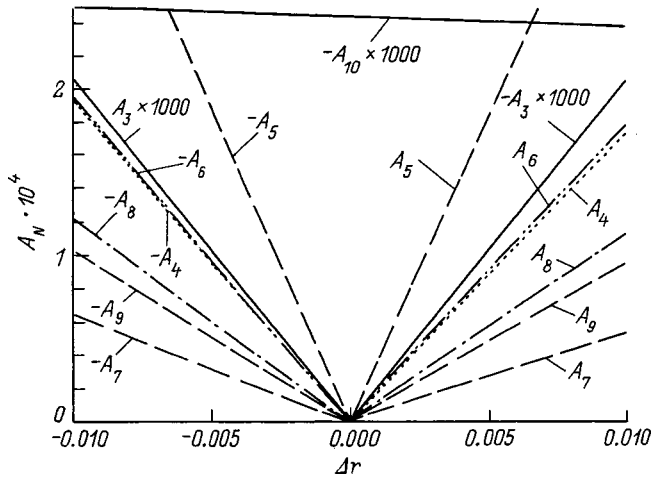

FIG. 4. Variations of the weighting coefficients A_N in the multipole expansion of the field with the radial displacement Δr of one rod.

the absence of a grounded electrode, this component is simply equal to a constant); thus, it is important to emphasize that our calculations of the spectral composition apply to the case $R=4r_0$.

In the case of a symmetric electrode geometry, the expansion of the field $\varphi_0(x,y)$ in terms of spatial harmonics has the form

$$\varphi_0(x,y)=A_0+\sum_{k=1}^{\infty} A_{4k}r^{4k}\cos(4k\varphi). \tag{7}$$

In particular, at the optimum ratio $r/r_0=1.145111$ for $R=4r_0$, the calculations yield

$$\varphi_0(x,y)=\mathrm{Re}\left(1.676057+0.000670\frac{z^4}{r_0^4}-0.000138\frac{z^8}{r_0^8}+0.000008\frac{z^{12}}{r_0^{12}}+0.0000005\frac{z^{16}}{r_0^{16}}\right), \tag{8}$$

where z is the complex coordinate of the field point.

As with the field $\varphi_2(x,y)$, the spectral composition of the field $\varphi_0(x,y)$ depends on the electrode geometry. Figure 3 shows the calculated amplitudes of the harmonics of the field $\varphi_0(x,y)$ as functions of the ratio r/r_0. This figure shows that an imbalance in the supply voltages has the greatest effect on the lowest quadrupole harmonic with $N=4$. The presence of a zero harmonic $N=0$ causes a small change in the transport energy of low mass analyzed ions and may be substantial for high mass ions.

3) *Linear displacements.* Figure 4 shows the amplitudes A_N of the spatial harmonics as functions of the radial displacement Δr of a single rod relative to its optimum position when $A_6=0$. The unit of length is taken to be the design parameter r_0. The signs of the coefficients A_N correspond to the case where a rod with a positive potential is shifted. A displacement of $\Delta r=0.01$ corresponds to an absolute drift of a rod by 50 μm for a standard electrode diameter of 10 mm. It can be seen that, besides the nonremoveable multipoles with $N=10, 14, 18, \ldots$, the displacement of a single electrode causes all the spectral components to appear, with the amplitudes of these harmonics being comparable in order of magnitude. The amplitudes of the odd, lowest multipoles with $N=3.5$ increase the most. Displacing an electrode inward into the analyzer ($\Delta r<0$) leads to a more rapid growth in the lowest even harmonics with $N=4$ and 6 than in the case of positive displacements. The amplitude of the decapole component A_{10} is 0.25% of the fundamental quadrupole component and varies little with the radial displacement of the electrode. The nonremoveable harmonics $N=14$ and 18 have amplitudes $A_{14}\approx 2.8\times 10^{-4}$ and $A_{18}\approx 2.2\times 10^{-5}$. The displacement of an electrode produces zeroth and first spatial harmonics of the field. The first harmonic ($N=1$) gives a shift in the axis for zero potential on the analyzer. Figure 4 shows that the small permissible radial displacements of an electrode such that the amplitudes of the lowest harmonics will be insignificant are $|\Delta r|<0.0003r_0$.

4) *Angular displacements.* The effect of an angular displacement Θ of a rod on the spectral composition of the spatial harmonics is illustrated in Fig. 5. Given the symmetry of the electrode positions, the field is identical for electrode displacements of $\pm\Theta$. The amplitudes A_N increase linearly with the angular displacement Θ, except for the hexagonal component $N=6$. Angular errors of 2° is the electrode placement yield powerful harmonics with $N=3$ and 4 and amplitudes of 1% of the fundamental. A_2. An angular displacement of 0.1° shifts the position of the electrode by $\Delta r = r_0\,\Theta=0.002$ and gives $A_3\approx 10^{-3}$. Thus, the angular positions of the electrodes should be maintained to within 0.001 rad.

5) *Nonidentical circular rods.* Figure 6 shows the effect of changing the cylinder radius on the formation of the multipole field components. The components shown here have amplitudes of the same order of magnitude. The permissible variations in the electrode radii which cause small distortions in the working field lie within $|\Delta r|<0.0005r_0$, and for

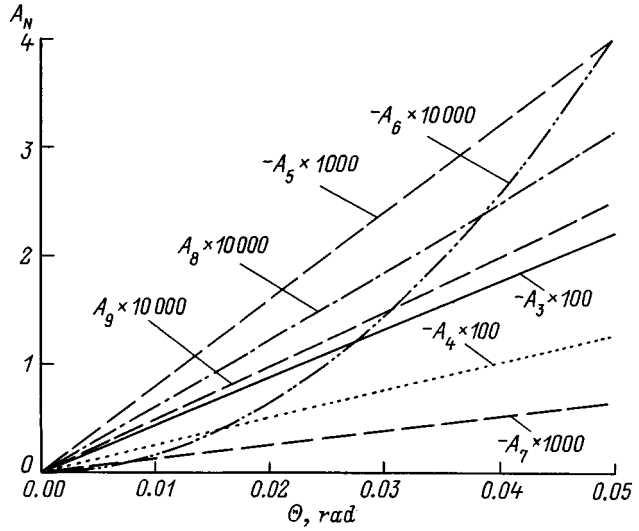

FIG. 5. The amplitudes A_N as functions of the angular displacement Θ of a single rod.

$r_0 = 5$ mm the cylinders must be machined with a tolerance of ± 2 μm.

A simple estimate of the effect of a machining error Δr in a quadrupole mass filter on the resolving power R gives[10]

$$R = \Delta r / 2 r_0, \tag{9}$$

and our results (Figs. 2–6) show that the maximum resolving power is related to the harmonics A_N by

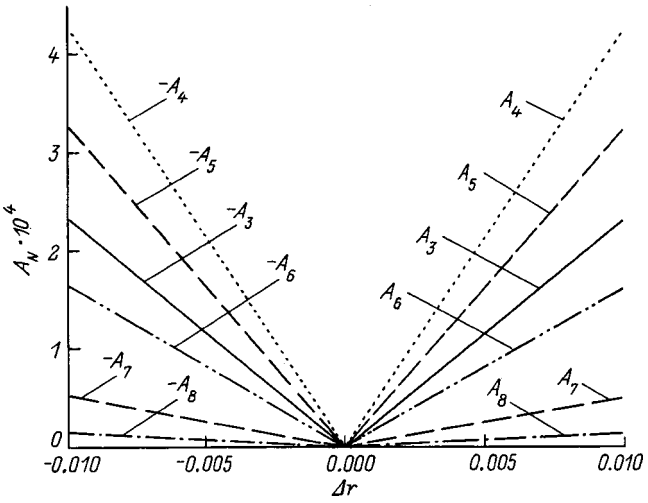

FIG. 6. The effect of a change Δr in the radius of a single rod on the spectral composition of the spatial harmonics A_N.

$$R = \text{const} \sum (1/A_N), \tag{10}$$

where const depends on the choice of level for the peak height from which the resolving power is determined, while its magnitude can be determined experimentally from the known spectral composition of the spatial harmonics.

CONCLUSION

We have examined the simplest types of displacements of circular quadrupole mass filter electrodes and calculated the multipole components in the expansion of the field for these cases. Displacements of 0–0.01 have little effect on the nonremoveable harmonics $N = 10, 14,$ and 18 but create strong multipole components in the lower orders $N = 3, 4, 5,$ and 6, which strongly distort the shape of a mass peak and limit the resolving power through a resonant enhancement in the amplitude of the oscillations of the ions in the multipole fields. An imbalance in the supply voltages ''turns on'' the spatial harmonics $N = 4, 8, 12, 16, \ldots$, of which the quadrupole moment, $N = 4$, is significant.

Knowledge of the spectral composition of the spatial harmonics of the field in a quadrupole mass filter with a controllable displacement of the electrodes will make it possible in the future to calculate the dips in the peaks and, thereby, to determine the allowable electrode drift and to estimate the effect of the quality of the field on the analyzer's throughput.

[1] C. Brunnee, Int. J. Mass Spectrom. Ion Processes **76**, 1–236 (1987).
[2] S. A. Lammert, Rapid Commun. Mass Spectrom. **11**, 821 (1997).
[3] F. von Bush and W. Paul, Z. Phys. **164**, 588 (1961).
[4] P. H. Dawson and N. R. Whetten, Int. J. Mass Spectrom. Ion Phys. **3**, 1 (1969).
[5] Y. Wang, J. Franzen, and K. P. Wanczek, Int. J. Mass Spectrom. Ion Processes **124**, 125 (1993).
[6] Y. Wang and J. Franzen, Int. J. Mass Spectrom. Ion Processes **132**, 155 (1994).
[7] N. V. Konenkov and S. S. Silakov, Zh. Tekh. Fiz. **61**(7), 146 (1991) [Sov. Phys. Tech. Phys. **36**, 807 (1991)].
[8] Ji-Feng Ying and D. J. Douglas, Rapid Commun. Mass Spectrom. **10**, 649 (1996).
[9] N. V. Konenkov and V. I. Kratenko, Int. J. Mass Spectrom. Ion Processes **108**, 115 (1991).
[10] G. M. Slobodenyuk, *Quadrupole Mass Spectrometers* [in Russian], Atomizdat, Moscow (1974).
[11] L. V. Danilov and E. S. Filippov (Eds.), *Computer Calculation of Electrical Circuit and Electromagnetic Fields*, Radio i Svyaz', Moscow (1983), 344 pp.
[12] A. J. Reuben, G. B. Smith, P. Moses *et al.*, Int. J. Mass Spectrom. Ion Processes **154**, 43 (1996).
[13] N. V. Konenkov, V. I. Kratenko, G. I. Mogil'chenko, and S. S. Silakov, Pis'ma Zh. Tekh. Fiz. **15**(15), 23 (1989) [Sov. Tech. Phys. Lett. **15**(8), 589 (1989)].

Translated by D. H. McNeill

Structural transitions and phase transitions in titanium thin films under irradiation by a nitrogen–hydrogen plasma

A. M. Chaplanov and E. N. Shcherbakova

Institute of Electronics, National Academy of Sciences of Belarus, 220841 Minsk, Belarus
(Submitted May 6, 1998; resubmitted September 9, 1998)
Zh. Tekh. Fiz. **69**, 102–108 (October 1999)

Structural and phase transitions in titanium films under nitrogen–hydrogen plasma treatment are investigated. Regularities in the formation and growth of titanium nitride are established as functions of the plasma irradiation parameters. The electrical resistivity of the irradiated films is determined. © *1999 American Institute of Physics.* [S1063-7842(99)01510-X]

In recent years a great deal of attention in thin-film technology has been devoted to developing new methods and improving existing methods of preparing films of refractory compounds. This is explained by the wide applications of such materials in microelectronics.[1] Nitrides are of special interest, due to the presence in them of a rare combination of properties: high hardness and melting temperature, chemical inertness, and high electrical and thermal conductivity. Thin films of nitrides of refractory metals are used to create Schottky diodes with a low potential barrier in fast integrated circuits, to passivate aluminum surfaces, and as barrier layers preventing aluminum diffusion into silicon, and in ohmic contacts.[2]

In the present work, using transmission electron microscopy and electron diffraction on a JEM-100CX electron microscope and an ÉMR-102 electron diffraction setup we have carried out a study of structural and phase transitions in titanium films under irradiation by a nitrogen–hydrogen plasma. In addition, we employed x-ray photoelectron spectroscopy (XPES) to determine the phase state of the films. The resistivity ρ of films deposited on glass-ceramic (Sitall) substrates and treated in an arc-discharge gas plasma was measured by the four-probe technique.[3]

Titanium films 100 nm in thickness were prepared by electron-beam evaporation in a 3×10^{-4} Pa vacuum onto freshly cleaved NaCl crystals and glass-ceramic substrates at a substrate temperature of 373 K. The films deposited on NaCl were separated from the substrate and placed on molybdenum reticules, the titanium–glass-ceramic compositions were scribed with the aim of obtaining samples with dimensions 10×10 mm for XPES studies and 10×5 mm for measurements of the resistivity. The titanium films on molybdenum reticules and on glass-ceramic substrates were then treated on a URMZ 279.026 setup in an arc-discharge gas plasma.

The pressure of the nitrogen–hydrogen mixture was kept at a level of 5–6 Pa for the starting level of vacuum in the working chamber equal to 1.33×10^{-2} Pa. The hydrogen content was 5% (Ref. 4). To determine the ion current of the plasma we used a Langmuir probe to which a reverse bias of 30 V was applied, which made it possible to measure the ionic component of the plasma current.[5] Under the conditions of the experiment ($P=5-6$ Pa, $U_{ca}=80$ A) the ion current density $J_i=4.0\pm0.2$ mA/cm^2. The radiation dose N_s was calculated using the formula[6]

$$N_s = J_i \cdot t/q, \tag{1}$$

where t is the irradiation time, q is the ion charge. N_s was varied within the limits $5\times10^{18}-3\times10^{19}$ cm^{-2} by varying the treatment time within the limits 3–20 min. The temperature on the surface of the films was measured with the help of a Chromel–Alumel thermocouple and was 400, 500, 550, 600, or 700 °C, depending on the cathode current.

Electron-microscopy and electron-diffraction studies showed that the deposited titanium films were polycrystalline, finely dispersed, and had an average grain size of 15–20 nm (Fig. 1a). To examine the kinetics of formation of titanium nitride as a function of temperature, we subjected the films to plasma treatment for 10 min at various temperatures. Calculation of the electron-diffraction patterns of films irradiated by a nitrogen–hydrogen plasma at 400 °C [Fig. 1b] indicate that the lattice period of the titanium films increased. This is due to diffusion and solution in the titanium of nitrogen, which leads to the formation of α-solid solution of nitrogen in titanium, which possesses a larger lattice period than titanium (Table I). As the temperature was increased to 500 °C, along with the rings belonging to the α-solid solution of nitrogen in titanium, lines appeared belonging to titanium nitride with a low content of Ti_2N [Fig. 1c, Table I]. Increasing the temperature further to 550 °C leads to the formation of titanium nitride TiN [Fig. 1d]. Films treated at this temperature consist of two phases: TiN and Ti_2N. Increasing the temperature to 600 °C leads to a complete rearrangement of the crystalline lattice of the films to a cubic lattice with $a=0.423\pm0.1$ nm, and the diffraction lines in the electron-diffraction pattern [Fig. 1e] can be interpreted as belonging to TiN. As the temperature was increased to 700 °C, the form of the electron-diffraction patterns did not change [Fig. 1f]. On the whole, the phase composition of the titanium films varies with temperature according to the following scheme:

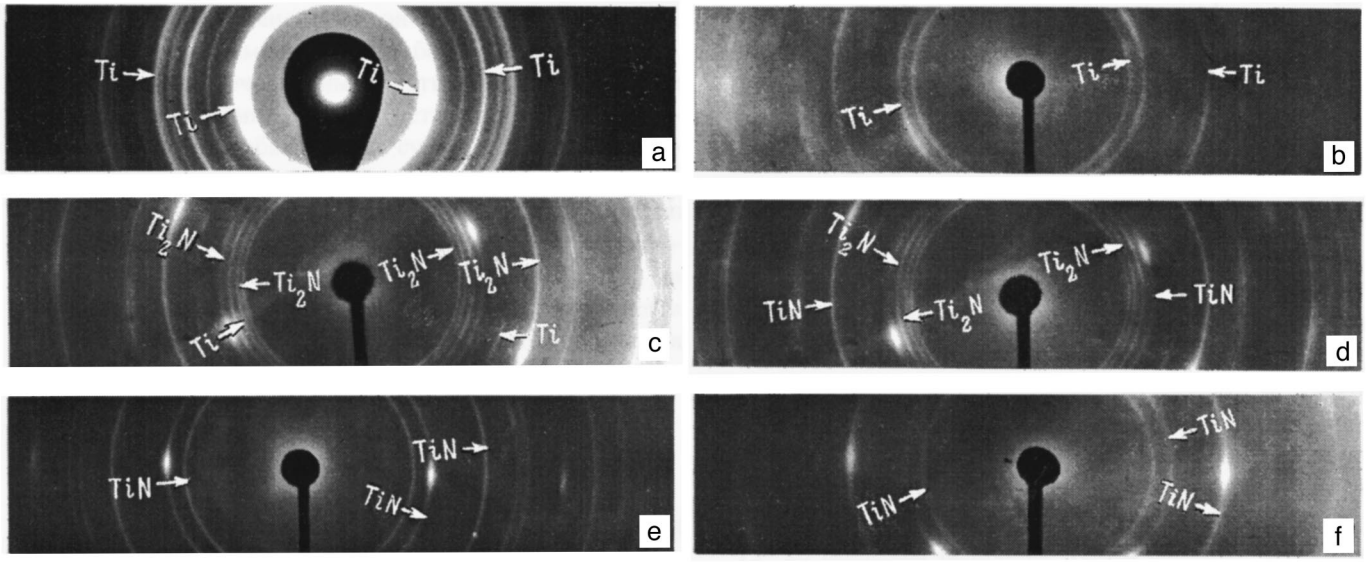

FIG. 1. Electron diffraction patterns of titanium films: a—initial sample, b–f—after irradiation by a nitrogen–hydrogen plasma at 400, 500, 550, 600, and 700 °C, respectively.

$$\mathrm{Ti} \xrightarrow{400\,^\circ\mathrm{C}} \alpha\text{-solid solution Ti} \xrightarrow{500\,^\circ\mathrm{C}} \alpha\text{-solid solution Ti}+\mathrm{Ti_2N} \xrightarrow{550\,^\circ\mathrm{C}} \mathrm{TiN}+\mathrm{Ti_2N} \xrightarrow{600-700\,^\circ\mathrm{C}} \mathrm{TiN}.$$

To examine trends in the formation of nitrides as functions of the ion dose, we irradiated the films at a constant temperature (700 °C) for 3, 5, 15, and 20 min. After treatment with an ion dose $N_s = 5\times10^{18}\,\mathrm{cm}^{-2}$ the diffraction rings in the electron-diffraction patterns of the titanium films belonging to titanium completely disappear, and the electron-diffraction pattern consists of rings pointing to the presence of TiN and Ti_2N [Fig. 2a]. When the radiation dose is increased to $7.5\times10^{18}\,\mathrm{cm}^{-2}$, the films become single-phase, consisting of pure TiN [Fig. 2b]. The films have a golden color characteristic of titanium nitride. Increasing the radiation dose further to 1.5×10^{19}, 2.3×10^{19}, and $3\times10^{19}\,\mathrm{cm}^{-2}$ has no effect on the phase composition of the films [Figs. 2c and 2d]. Results of deciphering the electron-diffraction patterns are summarized in Table II. On the

TABLE I. Dependence of the phase state of titanium films irradiated with a nitrogen-hydrogen plasma on the treatment temperature.

			T, °C				
d, nm	hkl	Starting films	400	500	550	700	800
0.261	010	···	Ti	···	···	···	···
0.259	101	···	···	Ti_2N	Ti_2N	···	···
0.256	010	Ti	···	···	···	···	···
0.247	200	···	···	Ti_2N	Ti_2N	···	···
0.244	111	···	···	···	···	TiN	TiN
0.229	111	···	···	···	Ti_2N	···	···
0.228	011	···	Ti	Ti	···	···	···
0.224	011	Ti	···	···	···	···	···
0.212	200	···	···	···	TiN	TiN	TiN
0.179	211	···	···	Ti_2N	Ti_2N	···	···
0.151	002	···	···	···	Ti_2N	···	···
0.1496	220	···	···	···	TiN	TiN	TiN
0.136	103	···	Ti	Ti	···	···	···
0.133	103	Ti	···	···	···	···	···
0.1277	311	···	···	···	TiN	TiN	TiN
0.126	201	···	Ti	···	···	···	···
0.125	213	···	···	Ti_2N	Ti_2N	···	···
0.1233	201	Ti	···	···	···	···	···
0.122	222	···	···	···	TiN	TiN	TiN
0.099	211	···	Ti	Ti···	···	···	
0.0946	211	Ti	···	···	···	···	···
0.094	420	···	···	···	TiN	TiN	TiN

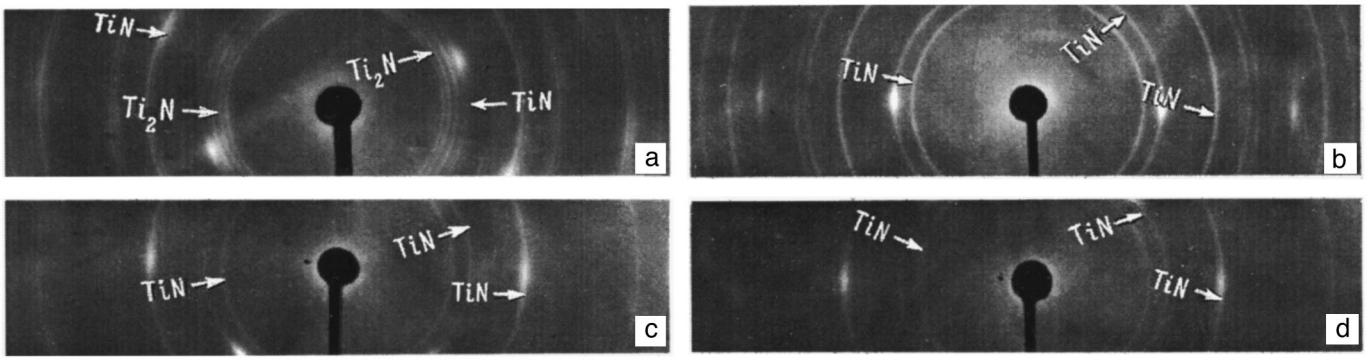

FIG. 2. Electron diffraction patterns of titanium films irradiated by a nitrogen–hydrogen plasma at 700 °C: $N_s = 7.5\times10^{18}$ (a), 7.5×10^{18} (b), 1.5×10^{19} (c), $3\times10^{19}\,\mathrm{cm}^{-2}$ (d).

whole, the variation of the phase composition of the titanium films as a function of radiation dose for irradiation at 700 °C can be represented as follows:

$$\mathrm{Ti} \xrightarrow{5\times10^{18}\,\mathrm{cm}^{-2}} \mathrm{TiN},\mathrm{Ti_2N} \xrightarrow{7.5\times10^{18}-3\times10^{19}\,\mathrm{cm}^{-2}} \mathrm{TiN}.$$

As the structural phase transition studies show, as the radiation dose is increased the titanium-nitride grains grow as a result of recrystallization. After irradiation with $N_s = 5\times10^{18}$ and $1.5\times10^{19}\,\mathrm{cm}^{-2}$ their mean diameter was 40 and 60 nm, respectively.

The XPES method was used to study the initial titanium films and films treated in a nitrogen–hydrogen plasma at 700 °C at radiation doses $N_s = 5\times10^{18}$ and $1.5\times10^{19}\,\mathrm{cm}^{-2}$. The C 1*s*, O 1*s*, Ti 2*p*, and N 1*s* spectra were taken on the surface of the films and after etching to a depth of 10 nm.

The oxygen spectra (Fig. 3a) correspond to the chemical state of oxygen in titanium dioxide TiO_2 with binding energy $E_c = 529.6$ eV in the initial film and binding energies 530.6 and 529.9 eV in films irradiated with doses equal to $N_s = 5\times10^{18}$ and $1.5\times10^{19}\,\mathrm{cm}^{-2}$, respectively. All these values of E_c agree with the values of the binding energy for oxygen in TiO_2. The oxygen peak intensities I_m differ substantially: the maximum value is equal to $I = 954$ arb. units in the initial films, and for the films irradiated by a nitrogen–hydrogen plasma the maximum value on the surface is equal to 468–492 arb. units; however, with depth it grows to 700 arb. units. The decrease of the oxygen content on the surface of the irradiated films is apparently explained by the action on the surface layer of hydrogen and nitrogen ions from the plasma, which knock out oxygen atoms from the surface of the titanium film. Varying the dose within the limits $5\times10^{18}-1.5\times10^{19}\,\mathrm{cm}^{-2}$ does not have a substantial quantitative effect on the oxygen content of the films.

On the basis of an analysis of the O 1*s* spectra on the surface after etching, it may be concluded that oxygen atoms adsorbed by the film during deposition are responsible for oxidizing processes in titanium films during plasma treatment while oxygen contained in the residual gases is compensated by the presence in the plasma of hydrogen ions and does not have an effect on the phase composition of the films.

The carbon spectra correspond to its free state in the graphite modification. The peak intensities in the irradiated films are decreased in comparison with the initial films (Fig. 3b). No significant difference in the spectra taken on the surface after etching is observed. The presence of carbon uniformly distributed with depth in the film is explained by its presence in the residual gases during deposition of the film due to the decomposition of diffusion-pump grease. The invariance of the C 1*s* spectra in the irradiated films indicates that interaction of titanium with carbon does not take place and that during treatment in a nitrogen–hydrogen plasma it does not have an effect on the phase composition of the films.

In the XPE spectra of Ti 2*p* (Fig. 3c) for the initial films, a peak is observed corresponding to the dioxide state of titanium ($E_c = 458.5$ eV). In the spectra of the irradiated films, in addition to this peak another peak is present corresponding to the state of titanium in TiN with binding energy 455.7 ± 0.2 eV. This again confirms that the phase detected in the irradiated films in the electron-microscopy studies belongs specifically to titanium nitride. The intensity of the dioxide peak is decreased in the treated films in comparison with the initial films, but grows with depth, which confirms our hypothesis, stated earlier in this paper, of a decrease in the oxygen content on the surface of the films due to the action of ions of the plasma.

For films irradiated with a dose equal to $5\times10^{18}\,\mathrm{cm}^{-2}$ the intensity and area of the peaks decreases somewhat with depth. This is evidence of nonuniform growth of titanium nitride under the given conditions of treatment. In films subjected to plasma treatment with a dose $N_s = 1.5\times10^{19}\,\mathrm{cm}^{-2}$ the differences in the TiN peaks in the spectra taken on the surface after etching are insignificant (Fig. 3c). Thus, after irradiation with a dose of $1.5\times10^{19}\,\mathrm{cm}^{-2}$ titanium nitride is uniformly distributed with depth.

In the N 1*s* spectra the principal peak corresponds to TiN for all the irradiated films, and its intensity is that much larger, the higher is the radiation dose (Fig. 3d).

In addition, in the spectra of films treated with a radiation dose $5\times10^{18}\,\mathrm{cm}^{-2}$, taken at some depth, a peak is present, corresponding to the α-solid solution of nitrogen in titanium. The presence in the spectra of some less intense

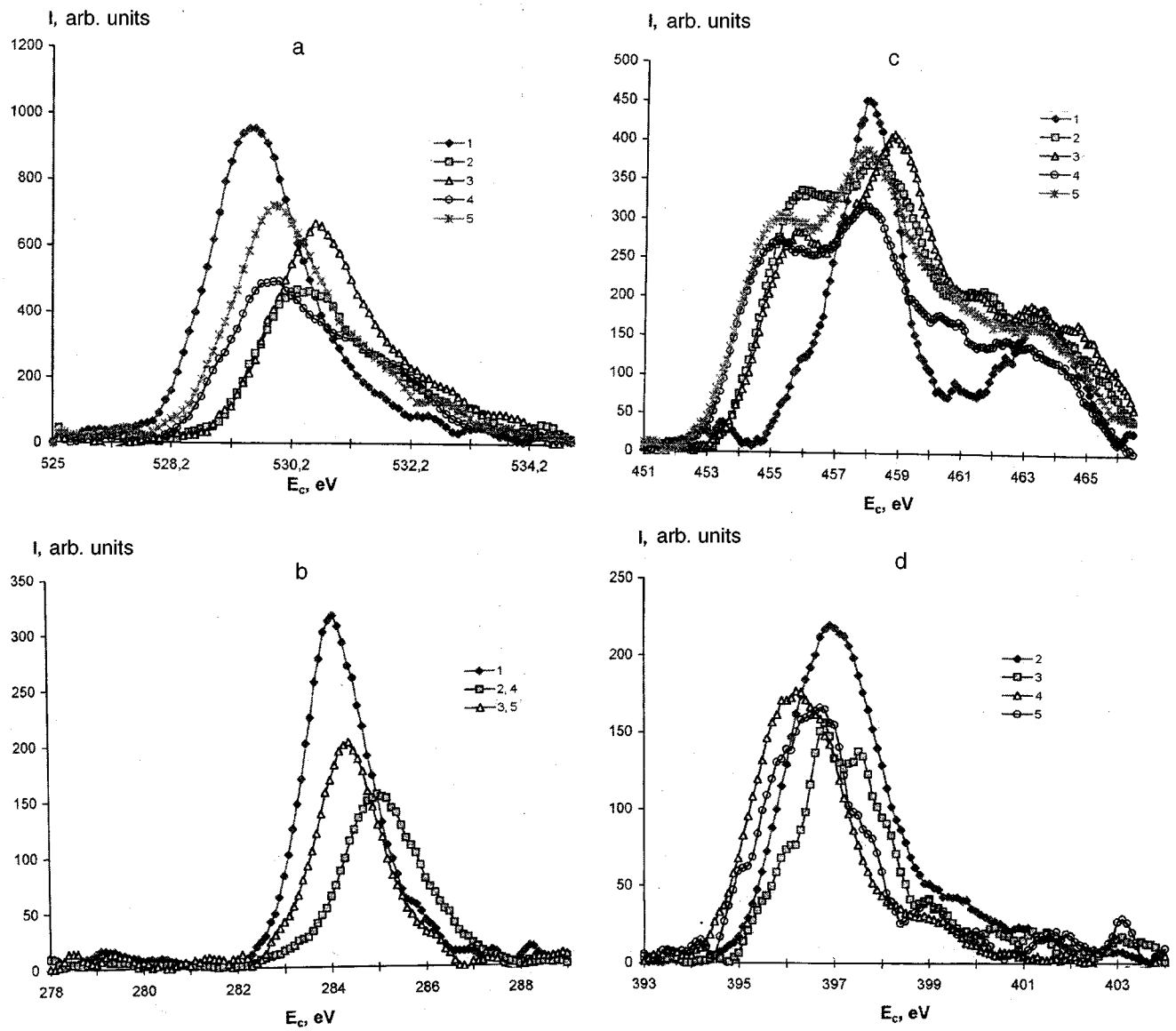

FIG. 3. X-ray photoelectron spectra: a—O 1*s*, b—C 1*s*, c—Ti 2*p*, d—N 1*s*; *1*—initial sample, *2*—after irradiation with $N_s = 5\times10^{18}\,\mathrm{cm}^{-2}$; *3*—after irradiation with $N_s = 5\times10^{18}\,\mathrm{cm}^{-2}$ and ion etching; *4*—after irradiation with $N_s = 1.5\times10^{19}\,\mathrm{cm}^{-2}$; *5*—after irradiation with $N_s = 1.5\times10^{19}\,\mathrm{cm}^{-2}$ and ion etching.

peaks is apparently explained by the formation in insignificant quantities of the titanium oxynitrides $TiN_{0.54}O_{0.17}$ ($E_c = 397.5$ eV), $TiN_{0.31}O_{0.44}$ ($E_c = 397.7$ eV), or compositions similar to them.[7]

Thus, the combined electron-microscopy and XPES studies confirm that when titanium films are irradiated by a nitrogen–hydrogen plasma, a nitride phase is formed in them having a face-centered cubic lattice. TiO_2 bonds are present in the irradiated and initial films, as indicated by the XPE spectra.

Our study of the processes of formation and growth of nitrides of refractory metals under the action of ions of a plasma shows that a nitrogen plasma contains several forms of the ions N^+, N_2^+, N_3^+, and N_4^+; however, it is ions of atomic nitrogen that are responsible for the formation of nitrides,[4] where the energy of an ion formed in an arc-discharge plasma exceeds the energy of the nitrogen atoms in dissociated ammonia by a factor of 3000 under conditions of ordinary nitrogenation. On the whole, the following processes take place during irradiation by a nitrogen–hydrogen plasma.

The nitrogen and hydrogen ions formed in the plasma move under the action of a magnetic field toward the surface of the film. The nitrogen ions actively interact with the surface of the titanium film and diffuse into it. Diffusion takes place preferentially on grain boundaries since the initial titanium film is finely dispersed. The ions and radicals arising as a result of molecular dissociation have an unpaired electron in their outer electron shell and as a consequence they exhibit a high degree of chemical activity. In addition, the concentration of the chemically active particles in the plasma, and also of the products of their reaction as a result of the presence of high-energy electrons substantially exceeds their concentration in thermodynamic equilibrium.[8] Thus, the (primarily atomic) nitrogen ions that have diffused into the film possess a high reactivity. Therefore, at 500 °C diffusion

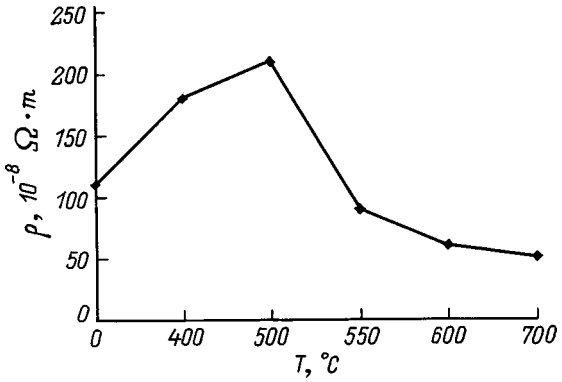

FIG. 4. Temperature dependence of the resistivity of titanium films irradiated by a nitrogen–hydrogen plasma.

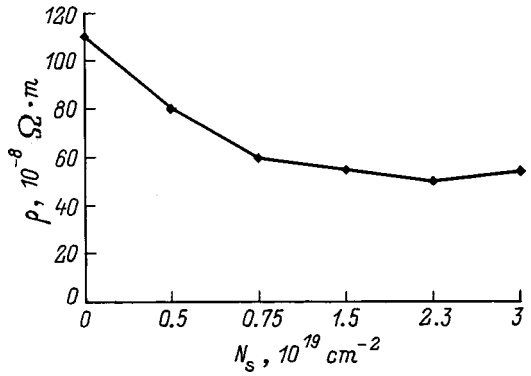

FIG. 5. Dependence of the resistivity of the titanium films on the radiation dose.

gives way to reaction–diffusion, and formation of a nitride phase with low nitrogen content—Ti_2N—commences in the films. Increasing the temperature to 600–700 °C and the radiation dose to $N_s = 7.5\times10^{18} - 3\times10^{19}\,\mathrm{cm}^{-2}$ leads to further saturation of the film with nitrogen, and its content in the film becomes sufficient for the formation of TiN, which has a wide interval of homogeneity from 30 to 53.7 at. % nitrogen.[9]

Increasing the radiation dose causes the growth of TiN grains as a consequence of processes of recrystallization of the films, which is connected with the thermal action of the plasma. The role of the hydrogen ions reduces to suppression of oxidation. Thanks to their presence, oxygen adsorbed by the film during deposition has no effect on the phase composition of the samples.[10]

The change in the phase composition of titanium films upon irradiation by a nitrogen–hydrogen plasma results in a change in the resistivity of the films. The initial films have a resistivity $\sim 110\times10^{-8}\,\Omega\cdot$m (Fig. 4). Plasma treatment at $T = 400$ °C results in a growth in the resistivity, which is connected with an increase in scattering of conduction electrons by implanted ions in the solid solution of nitrogen in titanium. As was established in Ref. 11, with increasing deviation from stoichiometry the mobility of the charge carriers decreases due to the presence of vacancies in the nitrogen sublattice which serve as scattering centers. Since an increase in the implanted impurity content by ~ 1 at. % leads to an increase in the resistivity of the titanium films by $8\times10^{-8}\,\Omega\cdot$m, it is possible to estimate the percent content of nitrogen in the titanium film after irradiation at 400 °C as 10 at. %. According to the phase diagram, this corresponds to an α-solid solution of nitrogen in titanium, which confirms our earlier analysis of the electron-diffraction patterns (Table I).

Increasing the temperature to 500 °C results in an increase in the concentration of the dissolved nitrogen to 14 at. % and, as a consequence, further growth of the resistivity to $\sim 210\times10^{-8}\,\Omega\cdot$m. The given percent value of nitrogen corresponds to the $\alpha+\varepsilon$ phases of titanium.

The abrupt decrease in the resistivity in the temperature interval 500–550 °C is due to the formation of TiN (Table I). Films prepared at 550 °C consist of two phases: $TiN+Ti_2N$; therefore the resistivity ρ of the titanium nitride films is $\sim 90\times10^{-8}\,\Omega\cdot$m. The decrease in the resistivity with the appearance of TiN is due to a phase transition from hexagonal Ti to cubic titanium nitride with a high nitrogen content, whose resistivity is higher than for titanium.[9]

Increasing the irradiation temperature to 600 °C leads to a further decrease of ρ to $60\times10^{-8}\,\Omega\cdot$m (Fig. 4) since the films become single-phase, consisting only of TiN. This phase is homogeneous over a wide interval of concentrations from 30 to 53.7 at. % nitrogen. With further increase of the temperature to 700 °C the resistivity varies only insignificantly, its decrease to $50\times10^{-8}\,\Omega\cdot$m is apparently due to an increase in the nitrogen content within the limits of the homogeneity interval of TiN, which results in the decrease in ρ.

The dependence of the resistivity on the irradiation dose for a given irradiation temperature is also determined by the phase state of the irradiated films (Fig. 5, Table II). For a dose of $5\times10^{18}\,\mathrm{cm}^{-2}$ the phases $TiN+Ti_2N$ are formed; therefore ρ is observed to decrease to $80\times10^{-8}\,\Omega\cdot$m, increasing the dose to $7.5\times10^{18}\,\mathrm{cm}^{-2}$ results in complete conversion to TiN and a further decrease in the resistivity to

TABLE II. Variation of the phase state of titanium films during treatment in a nitrogen hydrogen plasma at $T = 700$ °C as a function of the radiation dose.

d, nm	hkl	Before processing	Radiation dose, cm^{-2}	
			5×10^{18}	$7.5\times10^{18}-3\times10^{19}$
0.259	101	···	Ti_2N	···
0.256	010	Ti	···	···
0.247	200	···	Ti_2N	···
0.244	111	···	···	TiN
0.229	111	···	Ti_2N	···
0.224	011	Ti	···	···
0.212	200	···	TiN	TiN
0.179	211	···	Ti_2N	···
0.151	002	···	Ti_2N	···
0.1496	220	···	TiN	TiN
0.133	103	Ti	···	···
0.1277	311	···	TiN	TiN
0.125	213	···	Ti_2N	···
0.1233	201	Ti	···	···
0.122	222	···	TiN	TiN
0.0946	211	Ti	···	···
0.094	420	···	TiN	TiN

$\sim 55\times 10^{-8}\,\Omega\cdot$m. The difference in the values of ρ in the interval of radiation doses $7.5\times 10^{18}-3\times 10^{19}\,\mathrm{cm}^{-2}$ without any change in the phase composition of the films (Table II) is due to a difference in the nitrogen content within the limits of the homogeneity interval of TiN and an increase in the grain sizes when the radiation dose is increased, as a consequence of recrystallization processes.

Thus, our measurements of the resistivity of the prepared films have shown that by irradiating titanium films with a nitrogen–hydrogen plasma at temperatures 600–700 °C and radiation doses in the range $7.5\times 10^{18}-3\times 10^{19}\,\mathrm{cm}^{-2}$ it is possible to form films of titanium nitride with resistivity in the range $(50-60)\times 10^{-8}\,\Omega\cdot$m.

As a result of our studies, we have established that irradiation with a nitrogen–hydrogen plasma stimulates the formation and growth of titanium-nitride nuclei in titanium films and an increase in the grain sizes as a consequence of recrystallization processes. The resistivity of the irradiated films depends substantially on their phase composition and is determined by the plasma-treatment parameters. The results obtained here are indicative of the potential of using an arc-discharge nitrogen–hydrogen plasma to form thin films of titanium nitride with prescribed values of their resistivity.

[1] L. A. Dvorina and A. S. Dranenko, *Micro-Electronics and Refractory Compounds* [in Russian], Kiev (1996), 48 pp.
[2] M. Wittmer, J. Vac. Sci. Technol. A **3**, 1797 (1985).
[3] V. L. Kon'kov and R. A. Rubtsov, Izv. Vuzov: Ser. Fizika, No. 2, 135 (1965).
[4] B. N. Arzamasov, *Chemical Heat Treatment of Metals in Activated Gaseous Media* [in Russian], Mashinostroenie, Moscow (1979), 224 pp.
[5] L. A. Artsimovich, *Elementary Plasma Physics* [Blaisdell, New York, 1965; Atomizdat, Moscow, 1969, 192 pp.].
[6] W. C. Till and J. T. Luxon, *Integrated Circuits: Materials, Devices, and Fabrication* [Prentice-Hall, Englewood Cliffs, N.J., 1982; Mir, Moscow, 1985, 501 pp.].
[7] M. V. Kuznetsov, Yu. V. Zhuravlev, and V. A. Gubanov, Poverkhnost', No. 2, 86 (1992).
[8] B. S. Danilov, Itogi Nauki Tekh. Ser. Élektronika (VINITI, Moscow) **19**, 121–151 (1987).
[9] S. S. Kiparisov and Yu. V. Levinskiĭ, *Nitrogenation of Refractory Metals* [in Russian], Metallurgiya, Moscow, 1972.
[10] A. M. Chaplanov and A. N. Shibko, Neorg. Mater. **29**, 1477 (1993).
[11] O. Gernstenberg, Ann. Phys. (Leipzig) **11**(7–8), 354 (1963).

Translated by Paul F. Schippnick

Influence of lubricant on the motion of a body in an electromagnetic railgun accelerator. II. Hydrodynamics of a conducting lubricant

E. Yu. Flegontova and V. S. Yuferev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted January 21, 1998)
Zh. Tekh. Fiz. **69**, 109–116 (October 1999)

The flow of a liquid conducting film in the gap between the rail and the metallic projectile (armature) in a railgun is considered. In contrast to the usual problems in the theory of the hydrodynamics of lubricants, the statement of the problem here takes into account such additional forces acting on the liquid as the ponderomotive force and the inertial force arising as a consequence of the accelerated motion of the armature. As a result, even in the case of a gap of constant width not only is the existence of flows with negative velocities possible, but also of flows with negative flow rates. An analytical solution of the problem is obtained with the convective terms neglected. The influence of the convective terms is estimated by integrating the lubrication equations numerically. Primary attention is devoted to a calculation of the lubricant flow rate, viscous dissipation in the liquid layer, and the pressure distribution along the gap. Calculations are performed for specific cases of linearly expanding and linearly narrowing gaps. [S1063-7842(99)01610-4]

INTRODUCTION

In Ref. 1 it was shown that the intentional introduction of a liquid conducting lubricant into the gap between the rail and the metal projectile (armature) of a railgun can substantially alleviate the velocity skin effect and thereby improve the parameters of the railgun. The present paper is a continuation of Ref. 1 and is dedicated to a study of the flow of a conducting lubricant in the gap between the rail and the accelerated body. In principle, the role of the lubricant can be played by the film of molten metal formed as a result of melting-off of the sliding surfaces. Similar problems in the theory of lubrication have received ample study (see, e.g., Refs. 2–6). However, in the case of a railgun the use of the melt-off of the contacting surfaces themselves as lubrication hardly solves the main problem of improving the electrical contact between the rail and the armature, since it is not possible to maintain a stable, contiguous liquid film between the armature and the rail as a result of melting of their surfaces. Therefore it has been suggested that an external liquid be used as the lubricant, introduced into the gap between the rail and the armature. Specific means of introducing the lubricant into the gap have not been considered. In principle, this can be done either by depositing a corresponding layer on the surface of the rail or feeding the lubricant from the armature.

As is well known, there are several lubrication regimes,[7] which can be classified as hydrodynamic, boundary, or mixed, and elastohydrodynamic. In the hydrodynamic regime the sliding surfaces are completely separated by the film of lubricant. In this case, the pressure in the film is insufficient to lead to significant deformation of the surfaces bounding the film. In this case the profile of the gap is assumed to be known and independent of the motion of the lubricant in the gap. In the case of boundary lubrication, partial contact of salients of the surface roughness takes place and a large quantity of lubricant accumulates in the hollows between the salients. Therefore, boundary lubrication is a combination of hydrodynamic and mechanical contacts between the moving surfaces. Finally, the elastohydrodynamic regime arises under conditions in which elastic deformation of the contacting bodies plays a substantial role in the hydrodynamic process of lubrication and in turn depends on it. From the point of view of providing good electrical contact between the armature and the rails and reducing the wear on the sliding surfaces, the regime of boundary lubrication is undesirable although it is completely possible that it will arise in the initial stage of acceleration, when the velocity of the body is still insufficient to maintain by viscosity a lubricant film of sufficient thickness. Hydrodynamic and elastohydrodynamic lubrication are equivalent from the point of view of the passage of an electric current. The elastohydrodynamic regime exists if the thickness of the lubricant layer is comparable with the deformation of the contacting bodies. It is possible that this form of lubrication will arise in rail guns when lubricant is introduced into the gap between the armature and the rail since the only form of loading on the gap from the armature side is stresses caused by deformation of the armature under the action of inertial forces and the magnetic pressure. However, in the present work we restrict the discussion to the hydrodynamic regime, where the profile of the gap is assumed to be known and independent of the motion of the liquid. Primary attention was given to calculation of the flow rate of the lubricant as a function of the shape of the gap and its width, and also to calculation of the viscous dissipation and the pressure distribution in the liquid film.

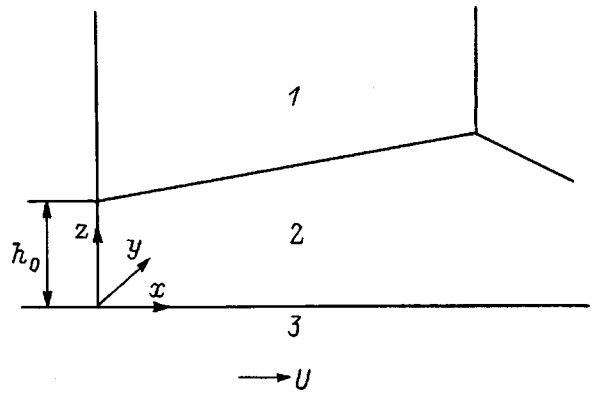

FIG. 1. Diagram of the gap: *1*—armature, *2*—film, *3*—rail.

STATEMENT OF THE PROBLEM

In contrast to the classical theory of hydrodynamic lubrication, in the modeling of the motion of a liquid film in a railgun it is necessary to take into account the presence of additional forces acting on the liquid. Such forces are the ponderomotive force and the inertial force associated with the accelerated motion of the armature. So far, in the treatment of such problems the indicated forces have not been taken into account.[8,9] A diagram of the gap is shown in Fig. 1. As usual, the coordinate system is bound to the moving armature. We consider the usual shape of the gap adopted in the solution of lubrication problems, where the lower boundary (in our case the rail surface) is planar, and the upper boundary (armature surface) can depend on the longitudinal coordinate. At the same time, under actual conditions, as was already mentioned, the shape of the gap can depend to a significant degree on the deformation of the armature and the rail and, consequently, the lower boundary can also be nonplanar. As in Ref. 1, we restrict the discussion to the planar problem, where only the y component of the magnetic induction B is nonzero. As a result, taking the usual assumptions of lubrication theory into account,[10] the motion of a conducting liquid in the gap will be described by the following system of equations:

$$\gamma\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+w\frac{\partial u}{\partial z}\right)=-\frac{\partial p_\Sigma}{\partial x}+\mu\frac{\partial^2 u}{\partial z^2}+\gamma a, \tag{1a}$$

$$\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}=0, \quad p_\Sigma=p_\Sigma(x) \tag{1b}$$

with boundary conditions

$$z=0,\ u=U; \quad z=h(x,t),\ u=w=0;$$

$$x=0,\ p_\Sigma=p_{in}; \quad x=l,\ p_\Sigma=p_M. \tag{2}$$

Here U is the velocity of the armature; u, w are the x and z components of the velocity of the lubricant; γ and μ are its density and dynamic viscosity; l is the length of the gap in the direction of motion, which in general differs from the length of the armature; a is the acceleration of the armature; p_Σ is the total pressure in the gap, equal to the sum of the magnetic $(B^2/2\mu_0)$ and hydrodynamic pressure (p); p_{in} is the pressure at the entrance to the gap; $p_M=p_a+B_0^2/2\mu_0$ is the pressure at the exit from the gap, here p_a is atmospheric pressure.

For the pressure p_M we used the estimate $p_M=L'I^2/2S$, where L' is the inductance of the rails per unit length, I is the current, and S is the cross-sectional area of the channel. The pressure p_{in} will, in general, differ from atmospheric pressure. In particular, if lubricant is deposited on the rail, this will occur due to compression of the lubricant as it enters the gap (see, e.g., Ref. 11). However, in the present study this difference is not taken into account. In the calculations we used a model lubricant with properties similar to those of liquid indium: $\gamma=7300\,\mathrm{kg/m^3}$ and $\mu=0.00146\,\mathrm{Pa\cdot s}$.

Problem (1) differs from the classical problem of the lubrication theory in that, first of all, due to the non-inertial character of the coordinate system, on the right-hand side of Eq. (1a) there appears an inertial term γa, and, second, the pressure that appears in the equation is the the total pressure, not the hydrodynamic pressure. As will be shown below, these two apparently minor differences lead to the appearance of solutions with qualitatively new properties.

Let us estimate the effect of the inertial terms in Eq. (1a). As a characteristic time scale we may choose either $t_1=h_*^2/\eta$, where h_* is the characteristic gap width and $\eta=\mu/\gamma$ is the kinematic viscosity, or $t_2=l/U$. During the indicated time the velocity of the armature varies by $\delta U=at_{1,2}$. As a result, the condition of quasistationary motion of the lubricant takes the form $\delta U/U<1$. Hence, setting $U=\sqrt{2aL}$, where L is the length of the channel, we obtain the following two conditions:

$$\frac{h_*^2}{\eta}\sqrt{\frac{a}{2L}}<1 \quad \text{and} \quad l/2L<1.$$

The second condition is always fulfilled, while the first condition, for typical values of the parameter $a=10^6\,\mathrm{m/s^2}$ and $L=1\,\mathrm{m}$, gives $h_*<1.7\times10^{-5}\,\mathrm{m}$. Thus, for films of thickness on the order of 10 μm or less, the time derivative in Eq. (1a) can be neglected and we can take the flow to be quasistationary. This is the approximation we will use below in the specific calculations. The contribution of the convective terms in Eq. (1a) is governed by the modified Reynolds number $\mathrm{Re}^*=Uh_*^2/\eta l$. For $U=2000\,\mathrm{m/s}$ and $l=0.01\,\mathrm{m}$ we find that $\mathrm{Re}^*<1$ for $h_*<1$ μm. Thus, the convective terms in Eq. (1a) can be neglected only in the case of sufficiently thin films.

The intensity of viscous dissipation in the film q_f and the flow rate of the lubricant Q are given by

$$q_f=\mu\int_0^h\left(\frac{\partial u}{\partial z}\right)^2 dz, \quad Q(x)=\int_0^h u\,dz, \tag{3}$$

and the thickness of the film of lubricant that remains on the rail is given by

$$h_{\mathrm{res}}=\frac{1}{U}\int_0^{z_*} u\,dz, \tag{4}$$

where z_* is the value of the z coordinate at which the velocity vanishes in the exit cross section of the gap.

The motion of the lubricant was calculated in two ways. In the case when the convective terms in Eq. (1a) can be neglected, we found an analytical solution of the problem, and to estimate the effect of the convective terms we integrated Eqs. (1) numerically.

ANALYTICAL SOLUTION OF PROBLEM (1)

If we neglect convective terms in Eq. (1a), then it is possible to obtain the distribution of the velocity of the lubricant over the cross section of the gap

$$u=\frac{1}{2\mu}\left(\frac{\partial p_\Sigma}{\partial x}-\gamma a\right)(z-h)z+U\left(1-\frac{z}{h}\right) \tag{5}$$

and also the Reynolds equation[12] for the pressure distribution along the gap

$$\frac{\partial}{\partial x}\left(h^3\frac{\partial p_\Sigma}{\partial x}\right)=3(2\mu U+\gamma a h^2)\frac{\partial h}{\partial x}. \tag{6}$$

It is not hard to see that Eq. (6) can be integrated by quadratures. Indeed, the first integration of Eq. (6) with respect to x gives

$$h^3\frac{\partial p_\Sigma}{\partial x}=(6\mu U+\gamma a h^2)h+c,$$

where c is the integration constant, determined by the boundary conditions (2).

After the second integration, taking the boundary conditions into account, we obtain the following expression for the pressure:

$$p_\Sigma(x)=p_\Sigma(0)+6\mu U\cdot F(x)+\gamma a x+cG(x), \tag{7a}$$

where

$$F(x)=\int_0^x\frac{d\tau}{h^2(\tau)},\quad G(x)=\int_0^x\frac{d\tau}{h^3(\tau)},$$

$$c=\frac{p_M-p_\Sigma(0)-\gamma a l-6\mu U\cdot F(l)}{G(l)}. \tag{7b}$$

We transform in Eqs. (5) and (7) to dimensionless variables

$$\tilde{x}=x/l,\quad \tilde{h}=h/h_0,\quad \tilde{u}=u/U,\quad \tilde{w}=w/U,$$

$$\tilde{p}_\Sigma=p_\Sigma/p_M,\quad \tilde{G}=Gh_0^3/l,$$

$$\tilde{F}=Fh_0^2/l,\quad \tilde{Q}=Q\frac{2}{Uh_0},$$

$$\tilde{q}_f=q_f\frac{h_0}{\mu U^2},\quad \tilde{h}_{\rm res}=\frac{h_{\rm res}}{h_0}, \tag{8}$$

where h_0 is the height of the entrance cross section of the gap (from here on we drop the tilde).

Then the pressure and the velocity profile in the gap in dimensionless units take the form

$$p_\Sigma(x)=p_\Sigma(0)+\beta\cdot(6F(x)+2\sigma x)+c_1\cdot G(x), \tag{9}$$

$$u=1-\frac{z}{h}+\left(\frac{1}{2\beta}\frac{\partial p_\Sigma}{\partial x}-\sigma\right)z(z-h), \tag{10}$$

where

$$\beta=\frac{\mu U l_f}{p_M h_0^2},\quad \sigma=\frac{\gamma a h_0^2}{2\mu U},$$

$$c_1=\frac{1-p_\Sigma(0)-2\beta\sigma-6\beta F(1)}{G(1)}. \tag{11a}$$

Expressing the acceleration of the body a in terms of the difference between the pressures $a=(p_M-p_{in})/\gamma_s l_{\rm eff}$, where $l_{\rm eff}$ and γ_s are the effective length of the armature and its density, we obtain for the coefficient σ

$$\sigma=\frac{\gamma(p_M-p_{in})h_0^2}{2\gamma_s l_{\rm eff}\cdot\mu U}. \tag{11b}$$

Differentiating expression (9) with respect to x and substituting it in Eq. (10), we find that

$$u=\left(1-\frac{z}{h}\right)\left(1-\left(3+\frac{c_1}{2\beta h}\right)\frac{z}{h}\right). \tag{12}$$

Now substituting expression (12) in Eqs. (2) and (3), we obtain expressions for the dimensionless flow rate and dimensionless viscous dissipation

$$Q=-\frac{c_1}{6\beta}=\frac{F(1)}{G(1)}+\frac{1}{6\beta G(1)}(2\sigma\beta-1+p_\Sigma(0)),$$

$$q_f=\frac{1}{h}\left(1+3\left(1-\frac{Q}{h}\right)^2\right).$$

Recalling expressions (11a) and (11b) for β and σ, we finally have

$$Q=\frac{F(1)}{G(1)}+\frac{1-p_\Sigma(0)}{6\beta G(1)}\left(\frac{l_f\gamma}{\gamma_s l_{\rm eff}}-1\right), \tag{13}$$

$$u=\left(1-\frac{z}{h}\right)\left(1-3\frac{z}{h}\left(1-\frac{Q}{h}\right)\right). \tag{14}$$

It can be seen from formulas (13) and (14) that the effect of the pressure drop in the gap and the acceleration of the body on the velocity distribution in the liquid film is governed by the parameter β. The latter in turn depends strongly on the thickness of the film. Indeed, for $U=2000$ m/s, $\gamma_s=2700$ kg/m^3, $l=l_{\rm eff}=0.01$ m, and $p_M=135$ MPa (the latter holds for $L'=3\times10^{-7}$ N/m, $I=300$ kA, and a channel cross section of 10^{-4} m^2) we find that $\beta=150$ for $h_0=1$ μm, and $\beta=1.5$ for $h_0=10$ μm. Thus, whereas in the first case the velocity distribution over the cross section of the film is governed mainly by viscous forces, in the second case inertial and ponderomotive forces also have a substantial effect.

Expression (14) shows that the velocity in some region of the cross section of the film can be negative. Indeed, the right-hand side of Eq. (14) becomes equal to zero when

$$z_*=h/3\left(1-\frac{Q}{h}\right).$$

Consequently, the condition for the appearance of negative velocities takes the form $z_*<h$ or $Q<2h/3$. Hence, substituting expression (14) into Eq. (4) we obtain an expression for the thickness of film remaining on the rail

$$h_{\text{res}}=\frac{Q}{2} \quad \text{for } z_* \geqslant h$$

and

$$h_{\text{res}}=h(1)\frac{9\left(1-\frac{Q}{h(1)}\right)-1}{54\left(1-\frac{Q}{h(1)}\right)^2} \quad \text{for } z_*<h.$$

It is also of interest to calculate the volume of lubricant that is used up as the armature passes through the channel of the railgun. We restrict our attention to the case where negative velocities do not arise in the exit cross section. Then the indicated volume in dimensional form is equal to

$$V=\frac{h_0}{2}\int_0^t UQdt.$$

In order to integrate analytically, we assume that the armature moves with uniform acceleration a, and the length of the rail is equal to L. Then

$$V=\frac{h_0}{2a}\int_0^{\sqrt{2La}} UQdU.$$

Substituting the expression for the flow rate (13) and expressions for β and a, we obtain

$$V=\frac{h_0 L}{2}\left(\frac{F(1)}{G(1)}+\frac{\sqrt{2}}{6}\left(\frac{l\gamma}{l_{\text{eff}}\gamma_s}-1\right)\frac{(1-p_\Sigma(0))^{1/2}}{G(1)}\times\frac{h_0^2}{l\mu}\left(p_M\gamma_s\frac{l_{\text{eff}}}{L}\right)^{1/2}\right). \tag{15}$$

ANALYSIS OF THE BEHAVIOR OF THE LUBRICANT IN THE GAP

As was shown in the preceding section, the velocity of the lubricant in the gap can become negative. This implies the appearance in the gap of return (circulational) flow. However, from the point of view of providing normal lubrication of the gap the appearance of negative velocities only at the entrance and exit cross sections of the gap is of fundamental importance. Indeed, this means that lubricant is partially drawn into the gap through the exit cross section and expelled from the gap through the entrance cross section. In turn, this means that if we wish, for example, to feed lubricant into the gap in front of the body, then the appearance of negative velocities in the exit cross section seems to indicate that lubricant will be sucked away from the rear edge of the armature.

Expression (13) shows that not only the velocity, but also the flow rate of the lubricant, can become negative. A necessary (but not sufficient) condition for the appearance of a negative flow rate, and also of negative velocities in the vicinity of the rear edge of the armature ($z=1$) is $l\gamma < l_{\text{eff}}\gamma_s$. From the physical point of view, this means that the lubricant per unit cross-sectional area is lighter than the armature and, consequently, will accelerate faster. In particular, for an aluminum armature and indium film the indicated condition is fulfilled if the effective length of the armature l_{eff} is 2.7 times greater than the length of the liquid film l. As can be seen from expression (13), decreasing the parameter β favors the appearance of negative flow rates. On the other hand, it follows from this same expression that for $l\gamma > l_{\text{eff}}\gamma_s$ as β decreases, the flow rate Q grows without limit, remaining positive. This leads to the appearance of a maximum in the velocity distribution (14) for $Q>4h/3$. The indicated maximum is unphysical since it implies that in the coordinate system bound to the rail, velocities appear directed opposite the motion of the body. The appearance of this maximum is probably due to neglecting the local time derivative in the equation of motion of the lubricant. Thus, we obtain another restriction on the applicability of the analytical solution.

The existence of solutions with negative flow rate gives reason to believe that lubricant can be injected into the gap not only in front of the body, but also through the rear cross section of the gap. However, to consider this question it is necessary to use more detailed models of a hydrodynamic lubricant which take account of the behavior of the liquid in the vicinity of the entrance and exit cross sections. It is also necessary to emphasize that in the classical theory of lubrication, due to the absence of ponderomotive forces, the second term in expression (13) is absent and therefore the above enumerated peculiarities of the solution generally do not arise.

It is not possible to analyze the behavior of the lubricant for an arbitrary shape of the gap. Therefore, in the remainder of this Section, we consider linear channels $h=1+kx$. In this case, expressions (7b) take the form $F(x)=x/1+kx$, $G(x)=(2x+kx^2)/(1+kx)^2$. Correspondingly, $F(1)=(1+k)^{-1}$, $G(1)=(k+2)(k+1)^{-2}/2$, and $h(1)=1+k$. As a result, for the flow rate of the lubricant we obtain

$$Q=\frac{2(1+k)}{2+k}+\frac{2(1+k)^2}{2+k}A, \tag{16}$$

where for convenience we have introduced the notation

$$A=\frac{1-p_\Sigma(0)}{6\beta}\left(\frac{l_f\gamma}{l_{\text{eff}}\gamma_s}-1\right).$$

It can be seen that for narrowing gaps, as the slope k is increased the flow rate decreases, and for expanding gaps it grows.

The condition for the appearance of negative velocities in the exit cross section of the gap takes the form

$$A<-(1-k)/3(1+k), \tag{17a}$$

and in the entrance cross section

$$A<-(1+2k)/3(1+k)^2. \tag{17b}$$

Hence for the thickness of the film remaining on the surface of the rail we obtain

$$h_{\text{res}}=\frac{1}{2+k}+A\frac{1+k}{2+k}, \quad \text{if } A>-\frac{1-k}{3(1+k)}$$

and

$$h_{\text{rex}}=(1+k)\frac{9\left(\frac{k}{2+k}-2A\frac{1+k}{2+k}\right)-1}{54\left(\frac{k}{2+k}-2A\frac{1+k}{2+k}\right)^2},\quad \text{if } A<-\frac{1-k}{3(1+k)}.$$

It is clear from relations (16) and (17) that in narrowing gaps increasing the slope hinders the appearance of negative velocities at the exit from the gap, and for expanding gaps—at the entrance to the gap. At the same time, for narrowing gaps negative velocities near the entrance cross section can also arise for positive values of A if $k<-0.5$. On the other hand, in the case of expanding gaps the analytical model does not give negative values at all while in the numerical solution (which is considered below) such velocities do appear.

In the particular case of a channel of constant cross section we have

$$Q=1+A,\quad h_{\text{res}}=\frac{1+A}{2},\quad \text{if } A>-\frac{1}{3}$$

and

$$h_{\text{res}}=-\frac{9A+1}{54A^2},\quad \text{if } A<-\frac{1}{3}. \tag{18}$$

Obviously, it would be desirable to have the thickness of the film remaining on the rail be as small as possible. It is clear from relation (18) that if $A\to-\infty$ then h_{res} will tend to zero. Thus, it would seem that we have the fundamental possibility to make the thickness of the remaining film as small as desired. However, this conclusion requires additional checking and refinement since it is based on a very approximate model.

And finally, for the volume of expended lubricant we obtain the expression

$$V=\frac{h_0L}{2}\frac{2(1+k)}{2+k}\left(1+(1+k)\frac{\sqrt{2}}{6}\left(\frac{l_f\gamma}{l_{\text{eff}}\gamma_s}-1\right)\times\frac{(1-p_\Sigma(0))^{1/2}}{G(1)}\frac{h_0^2}{l_f\mu}\left(p_M\gamma_s\frac{l_{\text{eff}}}{L}\right)^{1/2}\right). \tag{19}$$

From the point of view of a better capture of lubricant into the gap between the armature and the rail it would be desirable to use a narrowing gap. Figure 2 plots the dependence of the lubricant flow rate and the volume of expended lubricant on the degree of narrowing of the gap. It can be seen that as the degree of narrowing of the gap is increased the flow rate decreases. In this case, if the rate of narrowing of the gap is not large $|k|<0.5$, then for $\beta>1$ the dimensionless flow rate will differ only slightly from unity. The volume of expended lubricant for small film thicknesses (on the order of several microns) is determined mainly by the first term in large parentheses in expression (19), i.e., it is directly proportional to the film thickness and essentially independent of the acceleration parameters. With increase of the film thickness the role of the second term grows rapidly and for small rates of narrowing the dependence of V on the width of the gap approaches a cubic dependence. Figure 3

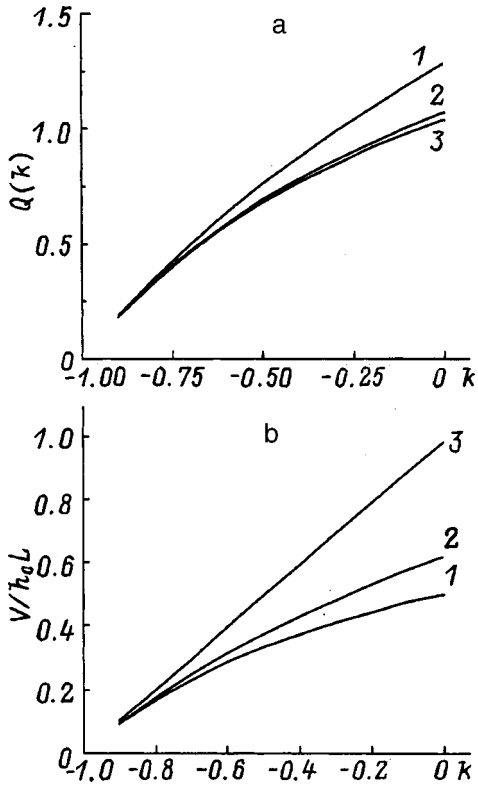

FIG. 2. Dependence of the flow rate of the lubricant (a) and the volume of expended lubricant (b) on the degree of narrowing of the gap, the parameter β, and the height of the entrance cross section of the gap h_0: a—$\beta=1$ (*1*), 4 (*2*), 10 (*3*); b—$h_0=10^{-6}$ (*1*), 10^{-5} (*2*), 2×10^{-5} m (*3*).

depicts the distribution of the viscous dissipation along the length of the gap for various values of k and β. It can again be seen that for $\beta>1$ and $|k|\leqslant0.5$ the dimensionless viscous dissipation is quite well approximated by the expression $q_f=1/h$, or in dimensional form $q_f=\mu U^2/h$, everywhere except in the immediate vicinity of the point $x=1$.

As was already pointed out, the pressure in the gap is composed of a hydrodynamic and a magnetic part. Here, since the magnetic field is concentrated in a narrow region about the point $x=1$ thanks to the velocity skin effect, the total pressure in the gap is equal to the hydrodynamic pressure everywhere except this region. As a result, even a layer of lubricant of constant thickness is capable of withstanding significant loads while in the case of ordinary hydrodynamic lubricant such a layer does not possess a carrying capability. The distribution of the total pressure along the length of the gap is depicted in Figs. 4 and 5. Figure 4 shows that for $\beta>1$ even in the case of quite moderate rates of narrowing the pressure inside the narrowing gap reaches large values, substantially exceeding not only the magnetic pressure applied to the accelerated body, but also the strength of the construction materials. This means that in such cases the hydrodynamic regime is apparently not realized and either boundary lubrication or elastohydrodynamic lubrication arises. Figure 5 depicts the pressure distribution in expanding gaps. In this case, the picture is the opposite and even for quite small rates of expansion the pressure in the film becomes negative when $\beta>1$. This means, first of all, that cavitation phenomena can arise in the film and, second, that

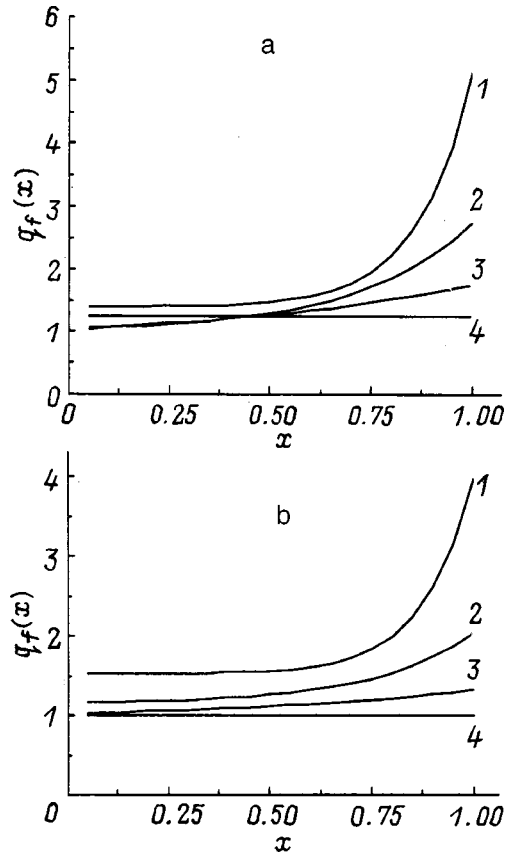

FIG. 3. Distribution of the viscous dissipation along the length of the gap: $\beta=1$ (a), 10 (b); $k=-0.6$ (*1*), -0.4 (*2*), -0.2 (*3*), 0.0 (*4*).

the model under consideration is probably not applicable for such rates of expansion of the gap.

NUMERICAL SOLUTION OF PROBLEM (1)

The derivation of the analytical solution rested on the assumption that the convective terms in Eq. (1) are small and can be neglected. However, in the presence of radical narrowing or expansion the indicated terms can become extremely important. We transform in Eq. (1) to the new variable $\eta=z/h(x,t)$. Using the dimensionless variables (8), we now have

$$\alpha\cdot\left(u\frac{\partial u}{\partial x}-\frac{1}{h}\frac{\partial}{\partial x}\left(h\int_0^{\eta}u\,d\eta\right)\frac{\partial u}{\partial\eta}\right)=\frac{\beta}{h^2}\frac{\partial^2 u}{\partial\eta^2}-\frac{\partial p_{\Sigma}}{\partial x}+2\sigma\beta, \tag{20}$$

for $\eta=0$ $u=1$, for $\eta=1$ $u=0$,

for $x=0$ $p_{\Sigma}=p_m/p_M$, for $x=1$ $p_{\Sigma}=1$, (21)

where $\alpha=\gamma U^2/p_M$.

Besides conditions (21) the solution of Eq. (20) should obey the condition of constancy of the flow rate, i.e., $h\int_0^1 u\,d\eta=Q$. The latter condition was used to calculate the gradient of the total pressure at each step of the integration with respect to x. At the entrance to the gap the velocity profile was assumed to be parabolic and was calculated according to formula (10). The value of $\partial p_{\Sigma}/\partial x$ at the entrance

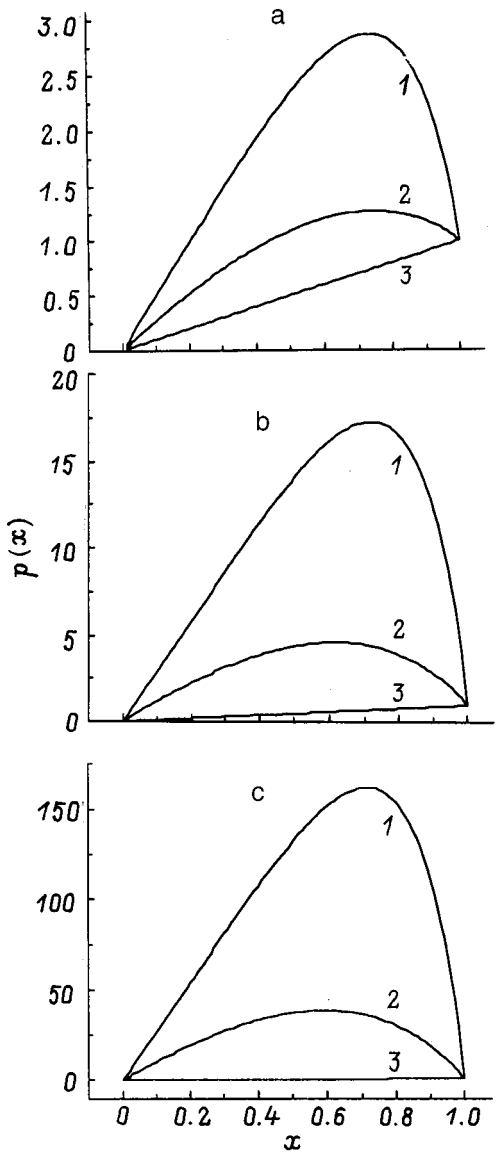

FIG. 4. Distribution of the pressure along a narrowing gap for different values of the parameter $\beta=1$ (a), 10 (b), 100 (c) and rates of narrowing $k=-0.6$ (*1*), -0.3 (*2*), and 0.0 (*3*).

to the gap at the start of the calculation was assigned arbitrarily (e.g., it was calculated according to an analytical formula obtained by solving an equation which neglected the convective terms). After the x coordinate reached the right boundary, the pressure at this point was compared with the boundary condition $p=p_M$ and a new value of $\partial p_{\Sigma}/\partial x$ at the entrance to the gap was found. The initial velocity profile, which had previously been assumed to be parabolic, was then recalculated and the calculations were repeated until the boundary condition $p=p_M$ at $x=1$ was satisfied with the prescribed degree of accuracy.

Equation (20) is an equation of parabolic type. Therefore, the problem under consideration is well-posed as long as the velocity u remains positive. The latter condition is violated as soon as such significant positive gradients appear in the gap that they will cause return flow, where the velocity of the lubricant is greater than the velocity of the body. In

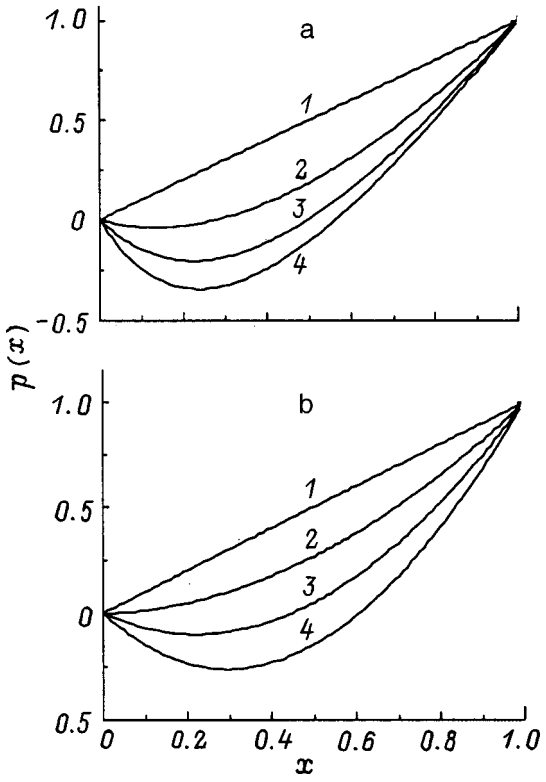

FIG. 5. Distribution of the pressure along an expanding gap for different values of the parameter $\beta=1$ (a) and 10 (b) and rates of expansion: a—$k=0.0$ (*1*), 0.3 (*2*), 0.6 (*3*), 0.9 (*4*); b—$k=0.0$ (*1*), 0.03 (*2*), 0.06 (*3*), 0.09 (*4*).

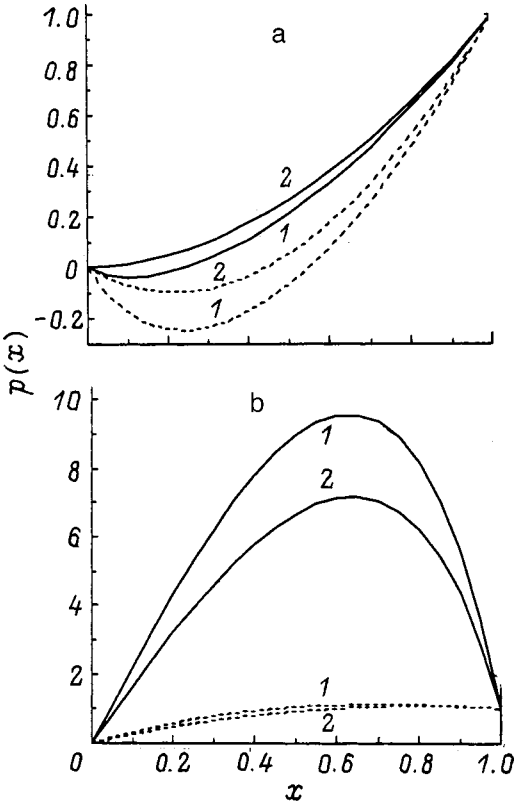

FIG. 6. Distribution of the pressure along an expanding (a) and a narrowing (b) gap: *1*—numerical solution, *2*—analytical solution; $\beta=10$, $\alpha=216$; a—$k=0.03$ (solid curve), 0.06 (dashed curve); b—$k=-0.4$ (solid curve), -0.5 (dashed curve).

this case, boundary conditions at the entrance to the gap are insufficient and it becomes necessary to use a boundary condition on the velocity at the exit from the gap.

It is clear from Eq. (20) that the role of the convective terms depends substantially on the magnitude of the parameter α. As a rule, $\alpha \gg 1$ as a consequence of the high velocities of the armature in a railgun. Therefore, it may be expected that the convective terms will have a considerable effect. Figure 6 depicts pressure distributions in linearly narrowing and a linearly expanding gaps, obtained numerically and by means of the analytical solution for a value of α corresponding to velocities on the order of 2–3 km/s and pressures in the range 100–300 MPa. It can be seen that the convective accelerations decrease the pressure in the expanding gaps and increase it in the narrowing gaps. Here the role of the convective terms grows with decreasing β and increasing k. Because of the appearance of return flows it is possible to obtain numerical solutions in a substantially narrower parameter region than is the case for analytical solutions.

CONCLUSION

The calculations performed here make it possible to explain a number of peculiarities of the motion of an electrically conducting lubricant in the gap between the rail and the accelerated body in a railgun. The results obtained are applicable to sufficiently thin films for relatively small degrees of narrowing or expansion of the gap. It appears that such films, because of the large velocities of the armature, are hardly capable of providing the regime of purely hydrodynamic lubrication and preventing direct contact between the rail and armature surfaces. In this case, we are dealing in all probability with the elastohydrodynamic regime. In the case significantly thicker films, lubricant flow in the gap becomes substantially nonstationary and possibly turbulent, and an explanation of its regularities requires further study. Such an interesting and important problem as the effect of the lubricant on the washing-out of molten metal from the gap between the rail and the armature upon melting of the latter under the action of Joule heating is also beyond the scope of the present study. The indicated problem has a direct bearing on the formation of a melting wave and the transition from current flow via a metallic contact to current flow via an arc contact. A study of this problem within the framework of the quasistationary model is impossible. Moreover, a numerical study of the given problem requires the use of the full set of Navier–Stokes equations, an account of the free surface of the film behind the armature, and a correct description of the flow of lubricant in the vicinity of the line of contact of three phases.

The authors wish to express their gratitude to É. M. Drobyshevskiĭ for proposing this topic of study and for helpful discussions during the completion of this work.

This work was carried out with the partial support of the

Defense Evaluation and Research Agency (DERA, Malvern, U.K.) (Contract No. SMC/4C2061).

[1] É. M. Drobyshevskiĭ, É. N. Kolesnikova, and V. S. Yuferev, Zh. Tekh. Fiz. **69**(7), 103 (1999) [Tech. Phys. **44**, 831 (1999)].
[2] W. R. D. Wilson, ASME J. Lubrication Techn. **98**(1), 22 (1976).
[3] A. K. Stiffler, Prob. Tren. **106**(3), 96 (1984).
[4] A. Bezhan, Sovremennoe Mashinostroenie. Ser. A, No. 9, pp. 1–9 (1989).
[5] O. E. Pushkarev and G. G. Kapustenko, Trenie Iznos. **13**, 611 (1992).
[6] V. A. Balakin and O. V. Pereverzeva, Trenie Iznos. **16**, 438 (1995).
[7] D. F. Moore, *Principles and Applications of Tribology* [Pergamon Press, Oxford, 1975; Mir, Moscow, 1978, 487 pp.].
[8] V. P. Bazilevskiĭ, R. M. Zayatdinov, and Yu. A. Kareev, in *Proceedings of the Second All-Union Seminar on the Dynamics of a High-Current Arc Discharge in a Magnetic Field* [in Russian], Novosibirsk (1991), pp. 285–303.
[9] B. A. Uryukov, A. D. Lebedev, and C. C. Milyaev, *Fourth European Symposium on Electromagnetic Launch Technology*, Gelle, Germany (1993). Paper N P1511.
[10] L. G. Loĭtsyanskiĭ, *Mechanics of Liquids and Gases* [in Russian], Nauka, Moscow, 1987, 441 pp.
[11] J. A. Tichy and S.-H. Chen, ASME J. Tribol. **107**, 32 (1985).
[12] O. Reynolds, Trans R. Soc. (London) **177**, Pt. 1, 157 (1996).

Translated by Paul F. Schippnick

Influence of lubricant on the motion of a body in an electromagnetic railgun accelerator. III. Temperature distribution in the armature, rail, and liquid film

É. M. Drobyshevskiĭ, É. N. Kolesnikova, and V. S. Yuferev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted January 21, 1998)
Zh. Tekh. Fiz. **69**, 117–125 (October 1999)

The effect of a conducting liquid lubricant on the heating of a rail and projectile (armature) is studied theoretically and experimentally. It is shown that both the Joule and friction heating of the accelerated body can be reduced significantly by using resistive liquid films. When the contact resistance of the film is high, its temperature is determined by two competing processes: Joule heating and heat removal by the moving film. As a result, the dependence of the film temperature on the magnetic Reynolds number, its thickness, and its resistance is nonmonotonic. In the limiting case where the velocity skin effect is completely suppressed and the magnetic Reynolds number is sufficiently high, the film temperature is extremely low. In intermediate cases, however, the film temperature can turn out to be rather high and exceed the melting point of the armature. Viscous dissipation in the liquid film has no significant effect on the temperature of the rail–armature interface until the melting of the armature is determined by Joule heating within it. In the case where the velocity skin effect is strongly suppressed, viscous dissipation along with Joule heating in the resistive film can become one of the major factors controlling the attainable velocity of bodies in railguns. © *1999 American Institute of Physics.* [S1063-7842(99)01710-9]

INTRODUCTION

This paper is the last of three devoted to a study of the effect of a conducting liquid film on various aspects of acceleration of a metal projectile (armature). The first paper[1] dealt with the influence of a lubricant on the distribution of the electrical current and magnetic field in a rail and armature and the second,[2] with the hydrodynamics of a conducting lubricant. The major purpose of this paper is to study the effect of a lubricant on the heating of the armature and rail. It is known that the sources of heat in a railgun are sliding friction and Joule dissipation. The problem of frictional heating in railgun accelerators has been studied previously.[3–6] Because dry friction is replaced by viscous dissipation in a liquid film when a lubricant is present, it is possible to greatly reduce this heating and prevent or, at least, delay the onset of melting of the armature (rail) at high sliding velocities. In addition, when highly resistive films are used, there is a substantial reduction in the velocity skin effect, which, in turn, reduces the Joule dissipation in the armature and delays the transition from metallic to arc contact between the rail and armature. These are all positive factors. However, there are some negative ones, as well. The higher resistivity of the film will cause additional Joule heating and, therefore, additional heating of the armature and the lower thermal conductivity of the lubricant will inhibit the removal of heat from the armature to the rail. Thus, the actual effect of a lubricant on the temperature distribution in the armature and rail cannot be determined without further, correct numerical calculations.

In studies of the heating of an armature, it is usually assumed that the temperature at a given point of the armature is determined by the local Joule heating at this point and heat conduction can be neglected. This approach has been used in all the papers devoted to the formation of a melting wave.[7–9] However, thin resistive films and coatings are present, it is no longer possible to neglect heat conduction and it must be taken into account in calculations of the temperature distribution in the neighborhood of these films (coatings). In this paper, we focus on the effect of a conducting lubricant on the temperature of the rail–armature interface and, in particular, on how the lubricant changes the onset time for melting at this interface.

STATEMENT OF THE PROBLEM

As before,[1,2] the problem was assumed to be two-dimensional and the armature shape was assumed to be rectangular. It was assumed that the velocity of the body and the current distribution are constant in time. Thus, we considered the heating of the rails and armature by constant Joule and viscous dissipation. The distribution of the electrical current density in the armature, rail, and film was taken from Ref. 1. In order to simplify the problem, it was assumed that all the thermal characteristics of the armature materials (aluminum), rails (copper), and liquid film (indium) are constant and independent of temperature. This assumption introduces an error of no more than 10–20%. The specific values of these characteristics are listed in Table I. The other parameters of the materials in the armature, rails, and film used in this paper are given in Refs. 1 and 2.

In a coordinate system attached to the armature, the temperature distribution in the armature, rails, and film obeys the following equations: in the armature

TABLE I. Thermal characteristics of the armature, rail, and lubricant.

Characteristics	Armature	Rail	Lubricant
Thermal conductivity	230 W/m·K	360 W/m·K	43 W/m·K
Specific heat	1000 J/kg·K	450 J/kg·K	245 J/kg·K
Thermal diffusivity	8.5×10^{-5} m^2/s	9.2×10^{-5} m^2/s	2.4×10^{-5} m^2/s

$$\gamma_a c_a \frac{\partial T_a}{\partial t} = \lambda_a \frac{\partial}{\partial x}\left(\frac{\partial T_a}{\partial x}\right) + \lambda_a \frac{\partial}{\partial z}\left(\frac{\partial T_a}{\partial z}\right) + \frac{j^2}{\sigma_a}, \tag{1}$$

for $x=0$, $T_a=T_{\text{room}}$; for $x=l$ $\partial T_a/\partial x=0$; for $z=h$ $T_a=T_f$,

$$\lambda_a \frac{\partial T_a}{\partial z} = \lambda_f \frac{\partial T_f}{\partial z};$$

for $z=d+h$, $\partial T_a/\partial z=0$; in the rail

$$\gamma_r c_r\left(\frac{\partial T_r}{\partial t} + U\frac{\partial T_r}{\partial x}\right) = \lambda_r \frac{\partial}{\partial x}\left(\frac{\partial T_r}{\partial x}\right) + \lambda_r \frac{\partial}{\partial z}\left(\frac{\partial T_r}{\partial z}\right) + \frac{j^2}{\sigma_r}, \tag{2}$$

for $x=0$, $T_r=T_{\text{room}}$; for $z=-\infty$, $T_r=T_{\text{room}}$; for $z=0$ and $0\leqslant x\leqslant l$, $T_a=T_f$,

$$\lambda_r \frac{\partial T_r}{\partial z} = \lambda_f \frac{\partial T_f}{\partial z};$$

and, in the liquid film

$$\gamma_f c_f\left(\frac{\partial T_f}{\partial t} + u\frac{\partial T_f}{\partial x} + w\frac{\partial T_f}{\partial z}\right) = \lambda_f \frac{\partial}{\partial z}\left(\frac{\partial T_f}{\partial z}\right) + \frac{j^2}{\sigma_f} + \mu_f\left(\frac{\partial u}{\partial z}\right)^2. \tag{3}$$

Here T is the temperature; the indices a, r, and f apply to the armature, rail, and film, respectively; j is the electrical current density; u and w are the components of the velocity of the lubricant; γ, c, λ, and σ are the density, specific heat, thermal conductivity, and electrical conductivity; μ_f is the dynamic viscosity of the lubricant; l is the length of the armature; d is half its height; and, h is the film thickness.

We shall assume that heat transfer across the film takes place sufficiently rapidly that the change in temperature over its thickness can be neglected. This assumption is valid if $\sqrt{a_f t_*} > h$, where a_f is the thermal diffusivity of the film and t_* is the characteristic time of the process. As will be seen from the following, this assumption was not always satisfied for the computational variants that were considered. Nevertheless, we have used this assumption since, first of all, it corresponds to the best conditions for heat exchange between the rail and the armature and, therefore, it can be used to find the minimum temperature of the armature surface in contact with the rail (film). Given these remarks, we integrate Eq. (3) perpendicular to the film, assuming that the temperature is constant along its thickness. We also approximate the viscous dissipation term in Eq. (3) by

$$\mu_f \int_0^h \left(\frac{\partial u}{\partial z}\right)^2 dz \approx \mu_f \frac{U^2}{h}.$$

Then we obtain

$$\lambda_a \frac{\partial T_a}{\partial z} - \lambda_r \frac{\partial T_r}{\partial z} = \gamma_f c_f\left(h\frac{\partial T_f}{\partial t} + Q\frac{\partial T_f}{\partial x}\right) - h\frac{j^2}{\sigma_f} - \mu_f \frac{U^2}{h}. \tag{4}$$

This equation, together with the condition $T_f=T_a=T_r$ at the contact interface between the rail and armature, serves as the boundary conditions for Eqs. (1) and (2). In this paper we did not consider heating of the body by dry friction, so that Eq. (4) contains only viscous dissipation. The latter increases without bound when the film thickness goes to zero. Thus, this model is applicable only to films that are not very thin, i.e. to those whose thicknesses exceed the dimensions of surface roughness of the rail and armature. At the same time, for calculating the temperature in the case where there is no liquid film, we have considered a model that neglects friction entirely.

REDUCTION TO DIMENSIONLESS FORM

We transform Eqs. (1), (2), and (4) to dimensionless form using the variables introduced in Refs. 1 and 2,

$$\tilde{B}=B/B_0, \quad \tilde{h}=h/h_0, \quad \tilde{x}=x/l,$$

$$\tilde{\mathbf{j}}=\tilde{\mathbf{j}}\,\frac{l}{\mu_0 B_0}, \quad \tilde{Q}=Q\frac{2}{Uh_0},$$

$\tilde{z}=z/l$ in the armature and $\tilde{z}=z\sqrt{\text{Re}}/l$ in the rail, as well as the variables

$$\tilde{T}=\frac{T-T_{\text{room}}}{T_m-T_{\text{room}}}, \quad \tilde{D}=D/D_a^*, \quad \tilde{t}=t/t_0,$$

where

$$t_0=\frac{l^2\mu_0\gamma_a c_a(T_m-T_{\text{room}})}{B_0^2 D_a^* |\tilde{\mathbf{j}}|^2_{\max}} \text{ and } \text{Re}=\frac{Ul}{D_r}. \tag{5}$$

Here Re is the magnetic Reynolds number; T_m is the armature temperature in degrees Celsius (660 °C for aluminum) and D_a^* is the diffusion coefficient for the magnetic field in the armature at the point where the current density is highest. The time scale t_0 was chosen so that the temperature of the armature at the point where the current density is highest will reach the melting point over a dimensionless time equal to 1 if the conductive flux is neglected. With this normalization, the onset time for melting of the armature will not exceed unity, regardless of the parameters of the problem. As a result, Eqs. (1), (2) and (4) take the form (the overhead tilde $\sim$ is omitted in the following): in the armature

$$\frac{\partial T_a}{\partial t} = \beta_T\left(\frac{\partial^2 T_a}{\partial x^2} + \frac{\partial^2 T}{\partial z^2}\right) + D_a J, \tag{6}$$

for $x=0$, $T_a=0$; for $x=1$, $\partial T_a/\partial x=0$; for $z=d$ $\partial T_a/\partial z=0$, where $a_a=\lambda_a/\gamma_a c_a$, and $J=|\tilde{\mathbf{j}}|/|\tilde{\mathbf{j}}|^2_{\max}$; in the rail

$$r_2\frac{\partial T_r}{\partial t} + \frac{\partial T_r}{\partial x} = r_3\frac{\partial^2 T_r}{\partial \eta^2} + r_1 J, \tag{7a}$$

for $x=0$, $T_r=0$; for $z=-\infty$, $T_r=0$, where

TABLE II. Maximum absolute value of the current density, melting onset time (in μs), and thickness of the thermal boundary layer for a uniform armature.

	Reynolds numbers, Re											
	30			100			300			1000		
Variants	$\lvert\tilde{\mathbf{j}}\rvert_{\max}$	t_0	$\delta_T\times10^3$	$\lvert\tilde{\mathbf{j}}\rvert_{\max}$	t_0	$\delta_T\times10^3$	$\lvert\tilde{\mathbf{j}}\rvert_{\max}$	t_0	$\delta_T\times10^3$	$\lvert\tilde{\mathbf{j}}\rvert_{\max}$	t_0	$\delta_t\times10^3$
1A	32	13.1	3.3	69.8	2.8	1.5	146	0.76	0.8	288	0.17	0.37
2A	28.9	16.2	3.7	58.5	3.9	1.8	113	1.1	0.92	234	0.26	0.45
3A	17.7	43.1	6.3	34.3	11.2	3.1	64.7	3.3	1.63	133	0.8	0.79
4A				20.3	32.8	5.3	35.4	1.1	3.1	65.2	3.3	1.7
5A				11.9	95.3	9.0	19.4	36.2	5.5	33.3	12.3	3
6A				3.9	889	27.4	6.1	362	17.5	11.3	105	9.4

$$r_1=\frac{1}{\beta_T\mathrm{Re}}\frac{a_r}{D_a^*}\frac{\lambda_a}{\lambda_r},\quad r_2=\frac{1}{\beta_T\mathrm{Re}}\frac{a_a}{D_r},$$

$$r_3=\frac{a_r}{D_r}\ \text{ and }\ a_r=\frac{\lambda_r}{\gamma_r c_r};\tag{7b}$$

and, at the rail–armature contact interface for $z=0$

$$\frac{\partial T_a}{\partial z}-s_5\frac{\partial T_r}{\partial z}=s_4 h\frac{\partial T_a}{\partial t}+s_3\frac{\partial T_a}{\partial x}-hs_1J-\frac{s_2}{h},$$

$$T_a=T_r,\tag{8a}$$

where

$$s_1=\frac{1}{\beta_T}\frac{D_f}{D_a^*}\frac{h_0}{l},\quad s_2=\mathrm{Re}^2\frac{D_r^2\mu}{h_0 l\lambda_a T_m},\quad s_3=\frac{\lambda_f}{\lambda_a}\frac{Q}{a_f},\tag{8b}$$

$$s_4=\frac{1}{\beta_T}\frac{a_a}{a_f}\frac{\lambda_f}{\lambda_a}\frac{h_0}{l},\quad s_5=\sqrt{\mathrm{Re}}\,\frac{\lambda_r}{\lambda_a},\tag{8c}$$

and for calculating J it was assumed that the x component of the current density in the film, averaged over the thickness, equals half the sum of the x components at the interface with the rail and armature, i.e.,

$$j_{x,\mathrm{film}}=(D_a j_{x,a}+D_r j_{x,r})/2D_f.\tag{8d}$$

It should be noted that Eqs. (5), (7b), and (8b) contain dimensional diffusion coefficients for the magnetic field. Recall, also, that in Eqs. (7a) and (8a), as well as in Eq. (6), J is equal to the ratio of the square of the absolute value of the current density to the maximum of the square of the absolute value of the current density in the armature.

TABLE III. Maximum absolute value of the current density, melting onset time (in μs) and thickness of the thermal boundary layer for an armature with a resistive layer. Re=1000.

Variants	$\lvert\tilde{\mathbf{j}}\rvert_{\max}$	t_0	$\delta_T\times10^3$
3B	33.5	1.2	1.0
4B	26.6	1.9	1.2
5B	19.5	3.8	1.7
6B	10.0	13.5	3.4

ANALYSIS OF EQS. (6)–(8)

The positive influence of a liquid film on the temperature at the rail–armature interface shows up primarily through the time scale t_0 and the coefficient β_T, which depend on the electric current density and, therefore, will increase as the velocity skin effect is reduced. The parameter β_T determines the thickness of the temperature boundary layer, δ_T, in the armature prior to the onset of melting, $\delta_T=\sqrt{\beta_T}$. In order to calculate t_0 and δ_T, it is necessary to specify the magnetic induction B_0. To calculate it, we have used the expression $B_0^2=\mu_0 L' I^2/S$, where L' is the linear inductance of the rails, I is the total current flowing through the armature, and S is the cross sectional area of the railgun channel. In the calculations, these quantities were assumed to be fixed, with $L'=3\times10^{-7}$ G/m, $I=300$ kA, and $S=10^{-4}$ m^{-2}. For the magnetic induction in the railgun channel this yields $B_0=18.4$ T. For calculating t_0 it is also necessary to specify the length l of the armature. In this section it was taken to be 0.01 m. Then, for a uniform armature and an armature with a resistive insert (variants A and C in Ref. 1), we obtain the following expressions for t_0 and δ_T:

$$t_0=13600/\lvert\tilde{\mathbf{j}}\rvert_{\max}^2\,\mu\mathrm{s},\quad \delta_T=0.105/\lvert\tilde{\mathbf{j}}\rvert_{\max}.$$

On the other hand, for an armature with a resistive layer (variant B in Ref. 1), the diffusion coefficient D_a^* Eq. (5) is 10 times greater. Thus, for this variant we shall have $t_0=1360/\lvert\tilde{\mathbf{j}}\rvert_{\max}^2\,\mu$s and $\delta_T=0.0333/\lvert\tilde{\mathbf{j}}\rvert_{\max}$.

Tables II–IV list the values of $\lvert\tilde{\mathbf{j}}\rvert_{\max}$, t_0, and δ_T for all the variants considered in Ref. 1. It is clear from these tables

TABLE IV. Maximum absolute value of the current density, melting onset time (in μs) and thickness of the thermal boundary layer for an armature with a chevron-type resistive insert in its rear portion.

	Reynolds numbers, Re					
	100			1000		
Variants	$\lvert\tilde{\mathbf{j}}\rvert_{\max}$	t_0	$\delta_t\times10^3$	$\lvert\tilde{\mathbf{j}}\rvert_{\max}$	t_0	$\delta_t\times10^3$
1C	33	12.4	3.2			
3C	21.2	30	5.0	122	0.9	0.88
4C	18	41.7	5.9	82	2.0	1.3
5C	14.2	67	7.5	52	5.0	2.1
6C	7.5	240	14.3	24	24	4.5

that the onset times for melting are extremely short and the thickness of the temperature boundary layer are extremely small. Even for a thick film (variant 6A) at Re=1000, this time is on the order of 100 μs, while the thickness of the boundary layer is 0.01 times the characteristic dimension of the armature. Nevertheless, in these variants the time for onset of melting of the armature increased by three orders of magnitude because of the presence of a film. Table III shows that using a resistive coating on the armature does not significantly reduce the melting onset time, since the effect of increasing the velocity skin depth mainly reduces to nothing because of the large increase in the resistance to the electrical current. A comparison of these tables with Tables II–IV of Ref. 1 shows that the maximum absolute value of the current density differs little from the maximum of the z component of the current density in all these variants, except those in series C which simulate an armature with an oblique, chevron-shaped rear wall. In this case, because a significant x component of the current develops, there is a rapid rise in the absolute value of the current density and, therefore, in the Joule heating. Thus, although the z component of the current density can be significantly reduced by this type of armature and, thereby, the velocity skin effect made smaller, again there is no real advantage in terms of reducing the heating of the armature (Table IV).

It should be noted that t_0 determines the onset time for melting under the assumption that the body is moving at a constant velocity. In a real situation, the time for onset of melting of the accelerated armature (as before, neglecting heat conduction) is determined by the condition

$$T_m - T_{\text{room}} = \frac{1}{c_a \gamma_a} \int_0^{t_0} \frac{j^2}{\sigma} dt.$$

Assuming uniformly accelerated motion of the body, we have

$$(T_m - T_{\text{room}}) \frac{\mu_0}{B_0^2} \frac{l^3 c_a \gamma_a}{D_a^*} \frac{a}{D_r} = \int_0^{\text{Re}^*} |\tilde{\mathbf{j}}|^2_{\max} d\text{Re}, \tag{9}$$

where $\text{Re}^* = a l t_0 / D_r$ and a is the acceleration of the body.

In the present case, $a = 5 \times 10^6$ m/s. As a result, Eq. (9) takes the form

$$\int_0^{\text{Re}^*} |\tilde{\mathbf{j}}|^2_{\max} d\text{Re} = 1.9 \times 10^4.$$

Using the data of Table II, we can obtain the following estimates for the magnetic Reynolds numbers at which melting starts: for variant 1A$\approx$30, for variant 3A$\approx$50$-$60, for variant 5A$\approx$150, and for variant 6A$\approx$400. Thus, only through lessening of the velocity skin effect is a resistive liquid film able to shift significantly the time the armature begins to melt.

TABLE V. Maximum absolute value of the current density in the rail at the boundary with the armature.

Variants	Re=300	Re=1000
1A	152	325
2A	156	338
3A	132	299
4A	101	224
5A	76	167
6A	38	102

Besides having a dependence on β_T and t_0, the temperature distribution in the rails and armature depends on the other coefficients which show up in Eqs. (7) and (8a). Substituting the characteristic values of the coefficients for the rail, armature, and film materials from Table I in Eqs. (7b) and (8c), we obtain

$$r_1 = 0.105 |\tilde{\mathbf{j}}|^2_{\max}/\text{Re}, \quad r_2 = 0.212 |\tilde{\mathbf{j}}|^2_{\max}/\text{Re},$$

$$r_3 = 0.263 \times 10^{-2}, \quad s_1 = 62.3 \varepsilon_1 D_f |\tilde{\mathbf{j}}|^2_{\max},$$

$$s_2 = 0.832 \times 10^{-7} \frac{\text{Re}^2}{\varepsilon_1}, \quad s_3 = 136 \varepsilon_1 \text{Re},$$

$$s_4 = 57.8 \varepsilon_1 |\tilde{\mathbf{j}}|^2_{\max}, \quad s_5 = 1.56 \sqrt{\text{Re}}, \tag{10}$$

where ε_1 and D_f are parameters introduced in Ref. 1. For Q and for s_3 in Eq. (8c) we have used the representation $Q = D_* \varepsilon_1 \text{Re}/2$ which follows from Eq. (11) of Ref. 1. The resulting expressions will be used below for analyzing heat transfer processes in the rails, film, and armature.

As noted in Ref. 1, the x component of the current in the rail is significantly higher than that in the armature. Thus, not surprisingly at first glance, the local Joule heating in the rail is greater than in the armature. This is evident from Table V, which lists the maximum absolute values of the current density in the rail at the interface with the armature, and Table VI, which lists values of the integral $\int_0^{0.9997} |\mathbf{j}|^2 dx$ calculated in the rail, film, and armature for $z=0$. It should be noted that, since the x component of the current in the film is much

TABLE VI. Value of $\int_0^{0.9997} |\mathbf{j}|^2 dx$ at $z=0$ in the rail, film, and armature.

	Re=300			Re=1000		
Variants	in the rail	in the film	in the armature	in the rail	in the film	in the armature
1A	70.2	⋯	46.3	190	⋯	121
2A	73.3	22.5	44.0	206	53.4	112
3A	79.6	18.8	34.6	240	44.0	83.8
4A	85.9	12.2	22.6	274	24.7	47.5
5A	92.4	7.7	13.8	301	14.0	26.1
6A	104	2.9	4.5	339	5.5	9.5

TABLE VII. Maximum temperature in the rail neglecting heat conduction.

Variants	Re=300	Re=1000
1A	0.025	0.020
2A	0.026	0.022
3A	0.028	0.025
4A	0.030	0.029
5A	0.033	0.032
6A	0.037	0.036

TABLE VIII. Maximum temperature in the film neglecting heat conduction.

Variants	Re=300	Re=1000
2A	0.248	0.177
3A	0.207	0.146
4A	0.560	0.341
5A	1.178	0.644
6A	0.443	0.254

lower than the z component, the absolute value of the current density in the film will be essentially equivalent to the z component of the current.

a) *Heat transfer in the rail.* The equation for heat transfer in the rail contains the hyperbolic operator

$$r_2\frac{\partial T_r}{\partial t}+\frac{\partial T_r}{\partial x}.$$

This means that without heat conduction, heat transfer in the rail will take place along the characteristics $t-r_2x=\text{const}$, i.e., the temperature at the cross section of the rail with coordinate x at time t will be determined by Joule heating in the region $x-(t/r_2)<x'<x$ lying upstream of the motion of the rail,

$$T_r=r_1\int_{\left(x-\frac{t}{r_2}\right)_+}^{x}J(x')dx', \tag{11}$$

where an expression of the form $(a)_+$ equals a, if $a>0$, and equals 0, if $a<0$. For T_r we can also write an equivalent expression,

$$T_r=\frac{r_1}{r_2}\int_0^t J(x')dt',$$

where

$$x'=\left(x-\frac{t-t'}{r_2}\right)_+.$$

Tables II–IV of Ref. 1 imply that even if a film is absent, when $r_2\gg 1$, the thickness of the velocity skin layer turns out to be substantially less than $1/r_2$. This means that the temperature distribution in the rail very quickly ceases to depend essentially on the time and in Eq. (11) we can set the lower limit equal to 0. Given Table VI and Eq. (10), it is easy to see that in all the variants considered here, the heating of the rail by its intrinsic Joule dissipation is negligible, as confirmed by the data of Table VII, which lists the maximum temperatures in the rail calculated according to Eq. (11). It is clear that as the contact resistance of the film increases, the temperature in the rail rises slowly, but remains low. Here it turns out that the temperature is essentially independent of the magnetic Reynolds number. This is explained by the fact that the as Re increases, the increased Joule heating is fully compensated by the heat carried away by the moving rail. Thus, as should be expected, the rail will have a cooling effect on the armature (or on the film), removing heat from its boundary with the rail. In the limiting case of a high contact resistance of the film ($D_f\to\infty$), when, as Ref. 1 implies, the square of the absolute value of the current density in the armature equals $|\tilde{\mathbf{j}}|_a^2=1$, while that in the rail equals $|\tilde{\mathbf{j}}|_r^2=1+(4/\pi)\mathrm{Re}x$, we find that the maximum temperature in the rail will be 0.067, regardless of the current or the armature velocity. Thus, even in this limit, the heat of the rail is negligible.

b) *Heat transfer in the film.* When highly resistive films are used, substantial Joule heating will occur right in the film. This effect can be estimated using the same analysis as done above for the rail. In fact, Eq. (8a) contains an analogous hyperbolic operator. Thus, if we again neglect thermal conductivity and viscous dissipation, then the temperature distribution in a film with constant thickness will be given by

$$T_{\text{film}}=\frac{s_1}{s_3}\int_{\left(x-\frac{s_3}{s_4}t\right)_+}J(x')dx. \tag{12}$$

As with the rail, the temperature distribution in the film rapidly becomes essentially independent of time and the lower limit in the integral can be set to zero. Table VIII shows the maximum values of the film temperature calculated using Eq. (12). Evidently, as the magnetic Reynolds number increases, the film temperature decreases; this is explained by the fact that as the velocity of the body increases, the rate of heat removal by the moving film increases more rapidly than the Joule heating within it. For Re=1000 these results correlate well with the numerical calculations of the temperature of the rail–armature interface presented in the next section. At the same time, for Re=300 the temperatures in Table VIII are twice as high and in one variant even exceed the melting point of the armature. Thus, as opposed to the rail, at Re=300, the motion of the film is insufficient to remove the heat that is generated and the excess heat will be delivered to the rail by heat conduction. However, this will occur only for films with a rather high thermal diffusivity, where the assumption of a constant temperature transverse to the film is valid. In the opposite case, Joule heating in the film at relatively low velocities of the body turns out to be the main reason for melting of the armature. It should also be noted that in the present calculations, we have assumed that the average film velocity is half that of the rail. For thick films, this assumption is too crude and the average velocity of these films will obviously differ less from that of the rail.

In the case of highly resistive films, therefore, their temperature can be comparable to that of the armature. Here the film temperature has a nonmonotonic dependence on its contact resistance. In this regard, it is again interesting to examine the limiting distribution of the magnetic field as $D_f\to\infty$. In this case, Eq. (8d) yields $|\tilde{\mathbf{j}}|_{\max}^2=1$ and $J=1+(D_r/D_f)$

$\times(\mathrm{Re}/\pi)x$. Given the estimate for D_f from Ref. 2, we can neglect the second term in J. As a result, the maximum temperature in the film at time $t=1$ will be $T_{\max,\mathrm{film}}=0.38D_f/\mathrm{Re}$. Substituting the estimate for D_f from Ref. 2 here, we obtain $T_{\max,\mathrm{film}}=0.19(1+3.1/\varepsilon_1\sqrt{\mathrm{Re}})$. This result is convincing enough, since, on one hand, as $\mathrm{Re}\to\infty$ the film temperature turns out to be well below that of the armature, while, on the other hand, as the magnetic Reynolds number decreases, the film temperature increases. This last point is consistent, in principle, with the data of Table VIII. Thus, the effect of a film on the heating of the armature can be nonmonotonic and have a maximum at a certain resistivity of the film and a certain finite Reynolds number. The following circumstance should also be noted. Since the absolute value of the current density in the film is determined by the z component of the current, in variants 2B–5B and 2C–5C, where this component of the current was significantly reduced by reducing resistive layers, we can expect the film temperature to be significantly lower.

c) *Effect of heat transfer from the armature to the rail on the temperature of the rail–armature interface.* This analysis has not taken the transfer of heat by conduction from the armature to the rail into account. In order to evaluate this process, we use the condition of balance between the heat fluxes at the rail–armature interface, which can be written in the form

$$\sqrt{\mathrm{Re}}\,\frac{T_{\mathrm{bound}}-T_{r,\infty}}{\delta_r}=\frac{\lambda_a}{\lambda_r}\,\frac{T_{a,\infty}-T_{\mathrm{bound}}}{\delta_T},$$

where $T_{r,\infty}$ and $T_{a,\infty}$ are the maximum temperatures at the boundary in the rail and in the armature calculated neglecting heat conduction, T_{bound} is the temperature of this interface, and δ_r is the thickness of the temperature layer in the rail, which equals $\sqrt{\delta_{VSL}a_r/D_r}$.

The above remarks imply that $T_{a,\infty}=1$, while $T_{r,\infty}=r_1\int_0^1 J(x',0)dx'$. As an estimate for δ_{VSL} we can either take the thickness of the velocity skin layer from Ref. 1 or use the estimate $\delta_{VSL}=\dfrac{1}{\max J}\int_0^1 J(x',0)dx'$.

The latter appears to be more accurate. As an example, let us consider variant 1A (without a film) with $\mathrm{Re}=1000$. In this case, $T_{r,\infty}=0.020$, while δ_{VSL} equals 0.0025. As a result, we find that, because of heat transfer from the armature to the rail, the temperature of the rail–armature interface turns out to be $T_{\mathrm{bound}}=0.117$, i.e., substantially lower than the melting point, so that melting of the armature begins at interior points, separated from the boundary by the thickness of the temperature boundary layer. This value of the temperature correlates fairly well with the numerical calculations presented below.

d) *Effect of viscous dissipation in the film on the temperature of the rail–armature interface.* The effect of viscous dissipation on the temperature of the rail–armature interface can estimated in a way similar to that of the previous section. Setting the dimensionless boundary layer thickness in the armature equal to $\delta_a=\sqrt{\pi\beta_T t}$ and in the rail, to $\delta_r=\sqrt{\pi a_r t/D_r r_2}$, using the condition (8A), and neglecting Joule heating yields the following expression for the temperature of the rail–armature interface:

$$T_{\mathrm{bound}}=\frac{2}{\sqrt{\pi}}\,\frac{s_2\sqrt{t}}{s_4\dfrac{2}{\sqrt{\pi}}\sqrt{t}+\dfrac{1}{\sqrt{\beta_T}}+\dfrac{\lambda_a}{\lambda_r}\sqrt{\dfrac{a_r}{D_r\mathrm{Re}r_2}}}. \tag{13}$$

As an example, let us consider variant 2A for $\mathrm{Re}=1000$. In this case, $r_2=11.6$, while $s_4=316$. As a result, for $t=1$ we obtain $T_{\mathrm{bound}}=0.156$, which, again, is in fair agreement with the numerical calculations. Thus, even in the case of thin films ($\varepsilon=10^{-4}$), viscous dissipation cannot significantly heat the rail–armature interface over the time prior to the onset of melting of the armature by Joule dissipation; therefore, introducing a liquid lubricant actually makes it possible to reduce the frictional heating substantially. Nevertheless, if the velocity skin effect is eliminated by employing a resistive coating on the rail and a liquid film is used only as a lubricant, then in the case of thin films, viscous dissipation will be one of the main processes determining the onset of armature melting. In fact, if we neglect the time derivative in Eq. (8a), then, given Eq. (10), it is easy to obtain the following expression for the case of uniformly accelerated motion in place of Eq. (13):

$$T_{\mathrm{bound}}=1.05\times10^{-7}\,\frac{a^{3/4}L^{5/4}la_a^{1/2}}{\varepsilon_1 D_r^2\left(1+\dfrac{\lambda_r}{\lambda_a}\sqrt{\dfrac{a_a}{a_r}}\right)},$$

where, recall, a is the acceleration of the body and L is the length of the railgun channel. For $a=5\times10^6\ \mathrm{m/s^2}$ and $L=1$ m. we find that if the dimensionless thickness of the liquid layer thickness ε_1 is less than 3×10^{-4}, then the temperature of the rail–armature interface reaches the melting point before the body leaves the accelerator channel. If, on the other hand, Joule heating in a resistive coating is taken into account, then it turns out that melting of the body can begin long before it leaves the railgun channel.

e) *Numerical solution of Eqs. (6)–(8).* Because the temperature boundary layer in the armature is so thin, in Eq. (6) we can neglect the second derivative with respect to the coordinate x. Here we shall seek the temperature distribution in the armature in the form

$$T_a=T_{a1}+T_{a2}, \tag{14}$$

where T_{a2} is determined by the Joule heating in the armature, $\partial T_{a2}/\partial t=D_aJ$, and, therefore, $T_{a2}=D_aJt$, while the heat conduction near the rail–armature interface, T_{a1}, obeys

$$\frac{\partial T_{a1}}{\partial t}=\beta_T\frac{\partial^2 T_{a1}}{\partial z_2}+\beta_T\frac{\partial^2 T_{a2}}{\partial z_2}. \tag{15}$$

The character of the distribution of Joule dissipation in the armature shows that the second term in Eq. (15) can be neglected. As a result, for calculating T_{a1} we obtain an ordinary heat conduction problem

$$\frac{\partial T_{a1}}{\partial t}=\beta_T\frac{\partial^2 T_{a1}}{\partial z_2} \tag{16}$$

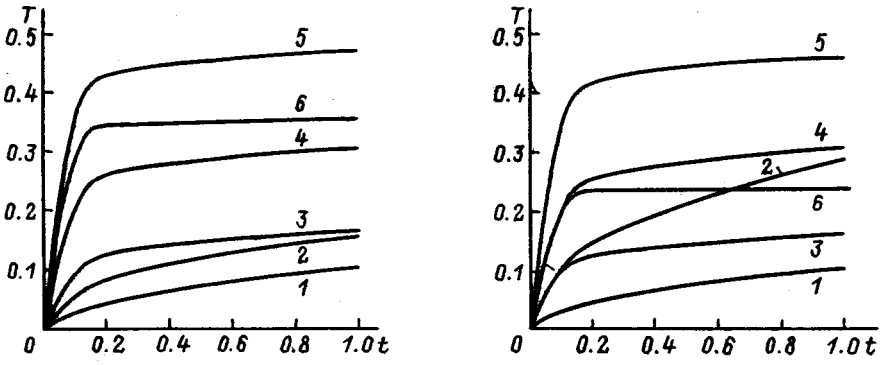

FIG. 1. Maximum temperature of the rail–armature interface as a function of time for a uniform armature: left Re=300; right — 1000; *1–6* — Variants 1A–6A, respectively.

with the conditions $T_{a1}=0$ for $t=0$ and $T_{a1}=0$ for $z=d$. Similarly, in the boundary conditions (8A) we can neglect the derivative with respect to z of T_{a2}. As a result, this condition takes the form

$$\frac{\partial T_{a1}}{\partial z}-s_5\frac{\partial T_r}{\partial z}=s_4h\left(\frac{\partial T_{a1}}{\partial t}+\frac{\partial T_{a2}}{\partial t}\right)+s_3\left(\frac{\partial T_{a1}}{\partial x}+\frac{\partial T_{a2}}{\partial x}\right)-hs_1J-\frac{s_2}{h}. \quad (17)$$

The system of Eqs. (7), (8b), and (14)–(17) is parabolic and a Crank–Nicholson scheme was used to solve it. The numerically calculated temperatures of the rail–armature interface are plotted in Figs. 1–3. These results are evidently in good agreement with the analysis in the previous sections. Thus, for Re=1000, beginning with variant 3A, the temperature of the rail–armature interface depends mainly on the Joule heating in the film itself, while the heat release to the armature is less than 20%. The temperature increase on going from variant 1A to variant 2A is related to the rather substantial viscous dissipation and the temperature decrease on going to variant 3A, to a reduction in the viscous dissipation owing to the increased thickness of the liquid film. The temperature drops on going from variant 5A to 6A because of a rise in the contact resistance of the film and the lessening of the velocity skin effect was caused by the increase in the film thickness rather than by the increase in its resistance. If, on the other hand, on going from variant 5A to variant 6A, the film thickness had been kept constant and its resistance increased by an equivalent number of times, then the maximum film temperature would have increased substantially. As Figs. 1 and 2 show, in variants 4–6 the temperature distribution at the interface actually reaches a stationary state fairly quickly. In all the cases considered here, the rail–armature interface temperature was substantially below the temperature inside the armature, since the Joule heating in the interior substantially exceeded the combined heat production at the boundary between the armature and rail. However, if the velocity skin effect is more completely suppressed, then the situation can change and the main source of heating of the armature becomes heat production in the resistive film and viscous dissipation. The interesting thing is that this phenomenon can arise in a rather early stage of the acceleration of the body, when its velocity is still relatively low. It is also interesting to note that in variants 3B–5B and 3C–5C, the temperature of the rail–armature interface was substantially lower than in variants 3A–5A. This is explained by the fact that the resistive layers in the armature made it possible to reduce the z component of the current density at the rail–armature interface significantly, which, in turn, led to a drop in the Joule heating in the film.

CONCLUSION

The above results yield the following conclusions.

Resistive liquid films can be used to reduce the Joule heating of the accelerated body substantially. Thus, in vari-

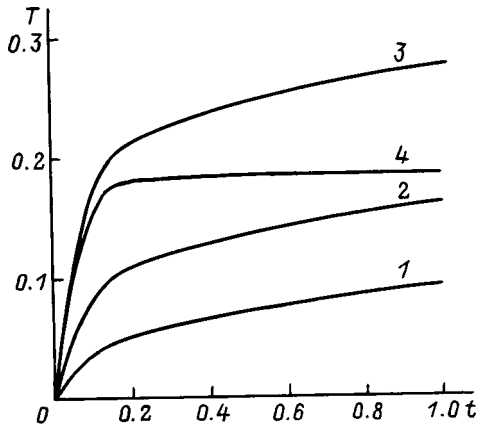

FIG. 2. Maximum temperature of the rail–armature interface as aa function of time for an armature with a chevron-type resistive insert: Re=1000; *1–4* — Variants 3C–6C, respectively.

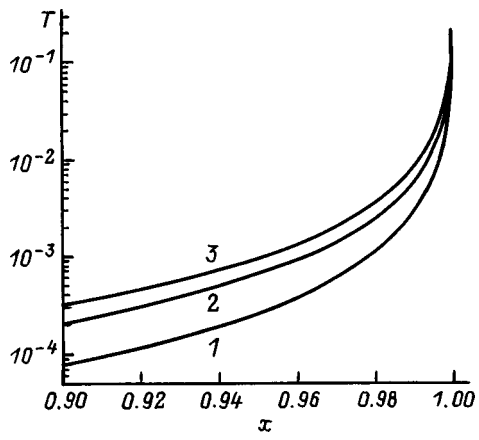

FIG. 3. Temperature distribution along the rail–armature interface at times $t=0.2$ (*1*), 0.6 (*2*), and 1.0 (*3*). Variant 1A (without a film); Re=1000.

ant 6A the melting onset time increased by almost three orders of magnitude compared to the case without a film. However, even here melting of the armature begins when the magnetic Reynolds number approaches 400, i.e., the velocity exceeds 1–1.5 km/s. Complete suppression of the velocity skin effect requires that the contact resistance of the film should exceed the resistance per unit area of the armature in the direction perpendicular to the film by tens of times.

Although Joule heating in the rail substantially exceeds that in the armature, the heating of the rail by its own Joule dissipation turns out to be negligible because the heat is carried away by the moving rail.

Applying a resistive coating to the armature does not significantly reduce the time for it to start melting, since the increase in the thickness of the velocity skin layer mainly reduces to nothing owing to the large increase in the resistance to the electrical current. Similarly, installing a chevron-type resistive insert also fails to significantly reduce the onset time for melting of the armature, since it causes a substantial rise in the x component of the current and, therefore, in the Joule dissipation within the armature. On the other hand, using these coatings and inserts reduces the z component of the current and, therefore, the Joule dissipation in the film, itself.

When the contact resistance of the film is high, its temperature is determined by Joule heating within the film, itself. The film temperature has a nonmonotonic dependence on the magnetic Reynolds number and the film thickness and resistance, since the temperature is determined by two competing processes: Joule heating and removal of heat by the moving film.

In the limiting case of complete suppression of the velocity skin effect and sufficiently high Re, the film temperature is extremely low. However, in intermediate cases, the film temperature can turn out to be extremely high and exceed the melting point of the armature.

Viscous dissipation does not have a significant effect on the rail–armature interface temperature in those cases where the melting of the armature is controlled by Joule heating within itself. When the velocity skin effect is strongly suppressed, viscous dissipation and Joule heating in the resistive film can become one of the main factors controlling the magnitude of the velocities attainable in a railgun.

This work was partially supported by DERA (Contract No. SMC/4C2061).

[1] É. M. Drobyshevskiĭ, É. N. Kolesnikova, and V. S. Yuferev, Zh. Tekh. Fiz. **69**(7), 103 (1999) [Tech. Phys. **44**, 831 (1999)].
[2] E. V. Flegontova and V. S. Yuferev, Zh. Tekh. Fiz. **69**(10), 109 (1999) [Tech. Phys. **44**, 1226 (1999)].
[3] B. A. Uryukov, A. D. Lebedev, and K. K. Milyaev, *Materials from the Second All-Union Seminar on the Dynamics of High-Current Arc Discharges in Magnetic Fields* [in Russian], Novosibirsk (1991), pp. 33–71.
[4] V. P. Bazilevskiĭ, R. M. Zayatdinov, and Yu. A. Kareev, *ibid.*, pp. 285–303.
[5] K. T. Hsich and B. K. Kim, IEEE Trans. Magn. **MAG-33**, 237 (1997).
[6] D. J. Hildenbrand, J. R. Rapka, and B. J. Long, IEEE Trans. Magn. **MAG-33**, 74 (1997).
[7] P. B. Parks, J. Appl. Phys. **67**, 3511 (1990).
[8] T. E. James and D. C. James, IEEE Trans. Magn. **MAG-31**, 162 (1995).
[9] L. C. Woods, IEEE Trans. Magn. **MAG-33**, 152 (1997).

Translated by D. H. McNeill

Interference detection of flaws in periodic objects in real time with adjustable sensitivity

A. M. Lyalikov and M. Yu. Serenko

Janko Kupala Grodno State University, 230023 Grodno, Belarus
(Submitted June 4, 1998)
Zh. Tekh. Fiz. **69**, 126–130 (October 1999)

A technique of interferometric testing for macroscopic defects in periodic objects is proposed which works in real time and with an adjustable measurement sensitivity. The possibility of compensating the aberrations of the interferometer by using a reference hologram is demonstrated. Interference patterns with different measurement sensitivity are presented which visualize flaws in a metal screen consisting of two overlapping grids. © *1999 American Institute of Physics.* [S1063-7842(99)01810-3]

INTRODUCTION

Defects (flaws) in objects having a periodic structure can be investigated by Moiré and holographic methods.[1–3] The image of a grid can be projected onto the surface of a reference object and onto the surface of the object to be tested. Then, by using double exposure and spatial filtering one obtains contour-difference Moiré fringes which visualize the surface defects.[1] The method of depositing a light-sensitive coating onto the surface of the object to be tested and then recording images of grids on it is also used.[2,4] To obtain contour lines of the surface relief in real time, two grids, arranged symmetrically about the line of sight, are projected simultaneously onto the investigated surface.[4] The possibility of rapidly adjusting the distance between the relief contours in the Moiré pattern in this scheme through the appropriate choice of the filtered diffraction orders of the diffraction gratings used for the projection has been investigated previously.[5]

There exists a class of objects with pronounced periodicity of their transmittance or surface texture. The contours of a surface with periodic texture are visualized in real time in the form of a pattern of Moiré fringes formed by the simultaneous imaging of an object with an amplitude grating.[6] Macroscopic surface defects of the slit mask of a television kinescope are visualized in real time by the Moiré method by superposing the image of the mask to be tested on the negative image of a reference mask.[3] In this latter work, to visualize surface defects with enhanced measurement sensitivity a holographic method was also employed, in which negatives of the reference and tested masks, mounted in optically conjugate planes in the optical analyzer, are used to obtain the interference pattern when symmetric orders of diffraction on the negatives are selected. It is also possible to use other holographic techniques. The deviation of the shape of the tested surface of the kinescope mask from that of the reference surface are visualized by overwriting the negatives of these surfaces with filtering of the complex-conjugate diffraction orders.[7] In both these works, to increase the measurement sensitivity in the visualization of defects of an object with periodic transmittance, pre-exposed negatives of the reference and tested objects were used. For arbitrary control of the width and orientation of the reference fringes, one of the negatives was pre-overwritten in a special scheme, which complicates the visualization process.

The present paper examines the possibility of visualizing defects of an object having a complex periodic structure of its transmittance by obtaining a two-beam interference pattern of one of the components of the complex object in one of the side diffraction orders.

EXPERIMENTAL SETUP

Figure 1a depicts the optical scheme of the experimental setup based on an IAB-451 shadow device. The basic scheme of the setup consists of a laser interferometer with superposition of the wide object beam and the narrow reference beam.[8] The setup allows one to form the interference pattern of an object with periodic transmittance using one of the side diffraction orders of the transmitted light and the regular reference beam, and also to form an image of the object and its spectrum. The narrow beam from the laser *1* is divided in the illuminator attachment *2* into two beams which are directed after reflection from the mirror *3* into the collimating objective *4, 5* of the shadow device. One of the beams, the unexpanded one, passes through the working zone along the edge of object *6* and is directed by the receiver objective *7, 8* and the mirror *9* into the receiver attachment *10*. The other beam is expanded into a wide parallel beam and illuminates the object *6*. The object is a transparency consisting of superimposed linear gratings. The optical scheme of the receiver attachment *10* is depicted in Fig. 1b. In the receiver attachment the narrow reference beak is expanded by the television system *14, 15* after reflection from mirrors *12, 13* and is directed by mirror *16* onto the beamsplitter *19*. The diaphragm *17* selects one of the beams diffracted by the object. The object beam selected by the diaphragm is combined with the reference beam at the beamsplitter *19*. The objective *18* has controllable travel and serves to equalize the curvature of the reference and object waves. The interference pattern is observed in real time in the recording plane *11* or is recorded on a photographic film mounted in the plane optically conjugate to the object being tested.

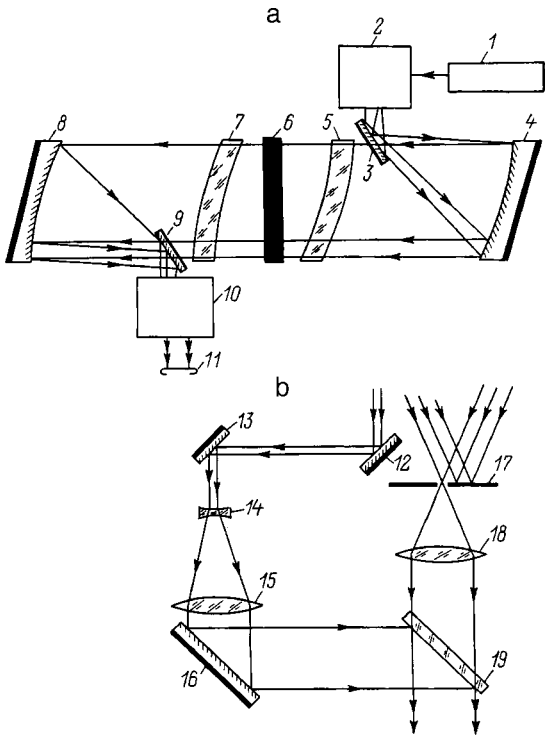

FIG. 1. Optical diagram of the experimental setup (a) and receiver attachment (b).

In general, the tested object *6* with periodic transmittance consists of a large number of superimposed linear gratings. When the object is illuminated by a parallel beam, from the entire, complicated spatial spectrum of the object the component corresponding to one of the side diffraction orders in one of the gratings is selected in the filtering plane *17*. The selected light wave, distorted by defects in the grating, is used to form in the recording plane *11* a two-beam interference pattern which visualizes defects in the object. The measurement sensitivity is regulated by choosing the appropriate diffraction order being selected in the filtering plane. The width and orientation of the interference fringes are regulated by tilting the mirrors *16* and *19* in the receiver attachment.

DESCRIPTION OF THE TECHNIQUE

Let us now consider the process of formation of an interference pattern, and also peculiarities and some possibilities of the method.

The amplitude transmittance of a composite transparency consisting of N superimposed amplitude gratings can be represented as

$$\tau(x,y)=\tau_1(x,y)\,\tau_2(x,y)\times\ldots\times\tau_N(x,y). \tag{1}$$

The xy coordinate system lies in the plane of the transparency, and the transmittance of each individual grating is described in general as a Fourier-series expansion[9]

$$\tau_k(x,y)=\sum_{-\infty}^{+\infty} a_{kn}\exp\{i[2\pi n(\xi_k x+\eta_k y)+n\varphi_k(x,y)]\}. \tag{2}$$

Here a_{kn} are the expansion coefficients; ξ_k and η_k are the spatial frequencies of the grating; and $\varphi_k(x,y)$ is the distortion of the periodic grating associated with defects in the object. The shape of the function $\varphi_k(x,y)$ is affected by local deviations of the shape of the surface of the object from planar and by distortions of the regularity of the spatial frequency of the grating.

When the transparency (1) is illuminated along its normal by a plane wave with complex amplitude a_0, the distribution of complex amplitudes of the diffracted waves has the form

$$A(x,y)=a_0\tau_1(x,y)\,\tau_2(x,y)\times\ldots\times\tau_N(x,y). \tag{3}$$

Expression (3) describes a set of waves diffracted to n orders into the angles α_k and β_k by each individual grating, where the direction cosines of the waves are equal to $\cos\alpha_{kn}=n\lambda\xi_k$ and $\cos\beta_{kn}=n\lambda\eta_k$, where λ is the wavelength of the illuminating wave. By virtue of the nonlinearity of expression (2), in addition to waves diffracted by individual scatterers, cross-diffraction will also be observed.[10]

Let us consider a wave that has been diffracted to the nth order by the kth grating,

$$A_{kn}(x,y)\sim a_{kn}\exp\left\{i\left[2\pi\left(\frac{\cos\alpha_{kn}}{\lambda}x+\frac{\cos\beta_{kn}}{\lambda}y\right)+n\varphi_k(x,y)\right]\right\}. \tag{4}$$

Here a_{kn} is the real amplitude of the wave. We assume that this wave is selected and interferes with a reference wave of the form

$$A_0(x,y)=a_0\exp\left[i2\pi\left(\frac{\cos\alpha_0}{\lambda}x+\frac{\cos\beta_0}{\lambda}y\right)\right], \tag{5}$$

where $\cos\alpha_0=\cos\alpha_{kn}$ and $\cos\beta_0=\cos\beta_{kn}$.

The illuminance distribution of the interference pattern is given by

$$I(x,y)\sim 1+\frac{2a_{kn}a_0}{a_{kn}^2+a_0^2}\cos[n\varphi_k(x',y')], \tag{6}$$

where $x'y'$ is the coordinate system in the recording plane.

The number of fringes in the image of the object for the interference pattern tuned to the infinitely wide fringe is equal, according to expression (6), to

$$C=\frac{n[\varphi_k(x',y')]_{\max}}{2\pi}, \tag{7}$$

where $[\varphi_k(x',y')]_{\max}$ is the maximum value over the area of the transparency of the quantity $\varphi_k(x,y)$ describing the deviation of the actual kth grating from the ideal.

It follows from formulas (6) and (7) that the measurement sensitivity is proportional to n. This makes it possible to control the measurement sensitivity directly when recording interference patterns by choosing the diffraction order selected by the diaphragm *17*. In the determination of the sign of $\varphi_k(x',y')$ from an interference pattern of the form (6) an indeterminacy arises which can be eliminated by recording an interference pattern tuned to finite fringes. For

such tuning the direction of curvature of the reference fringes is visualized unambiguously in both directions.

To tune the interference pattern to finite fringes, the tilt of the reference wave is varied as the interference pattern is being recorded. Referring to Eq. (5), let the tilt of the reference wave be defined by the small angles $\Delta\alpha$ and $\Delta\beta$

$$A_0(x,y)=a_0\exp\left\{i2\pi\left[\frac{\cos(\alpha_0+\Delta\alpha)}{\lambda}x+\frac{\cos(\beta_0+\Delta\beta)}{\lambda}y\right]\right\}. \tag{8}$$

Here the illuminance distribution of the interference pattern is given by

$$I(x',y')\sim 1+\frac{2aa_0}{a^2+a_0^2}\times\cos\left[2\pi\left(\frac{\Delta\alpha\sin\alpha_0}{\lambda}x'+\frac{\Delta\beta\sin\beta_0}{\lambda}y'\right)+n\varphi_k(x',y')\right]. \tag{9}$$

The direction of the reference fringes is determined by the relative values of $\Delta\alpha\sin\alpha_0$ and $\Delta\sin\beta_0$, and the period of the fringes is given by

$$T=\frac{\lambda}{\sqrt{(\Delta\alpha)^2\sin^2\alpha_0+(\Delta\beta)^2\sin^2\beta_0}}. \tag{10}$$

ABERRATIONS OF THE INTERFEROMETER AND THEIR COMPENSATION

Let us estimate the influence of aberrations of the interferometer on visualization of defects of periodic objects. The approach employed in a holographic interferometer may serve as the criterion of our estimate.[11]

Taking aberrations of the object arm of the interferometer into account, the wave diffracted into the nth order by the kth grating of the object *6* is written as

$$A_{kn}(x,y)\sim a_{kn}\exp\left\{i\left[\left(\frac{\cos\alpha_{kn}}{\lambda}x+\frac{\cos\beta_{kn}}{\lambda}y\right)+n\varphi_k(x,y)+\psi(x,y)\right]\right\}, \tag{11}$$

where $\psi(x,y)$ are the phase distortions of the wave due to the presence of aberrations.

The illuminance distribution of the interference pattern now has the form

$$I(x',y')\sim 1+\frac{a_{kn}a_0}{a_{kn}^2+a_0^2}\cos[n\varphi_k(x',y')+\psi(x',y')]. \tag{12}$$

As follows from Eq. (12), in the proposed method of visualizing defects in periodic objects the ratio of the useful signal $n\varphi_k(x',y')$ to the distortions $\psi(x',y')$ is proportional to the number n of the implemented diffraction order. As the number n of the selected order is increased, with the aim of enhancing the measurement sensitivity, the influence of aberrations of the interferometer decreases. In interferometry it is assumed that shifts of the interference fringe by less than 0.1 period lie within the limits of measurement error.[12,13] Consequently, we will assume that in the proposed approach aberrations of the object arm of the interferometer not exceeding 0.2λ can be neglected when visualizing defects of periodic objects. Aberrations of greater magnitude must be compensated.

In the holographic interferometry of phase objects the compensation of aberrations of the interferometer is implemented by methods employing reference holograms.[11] The essence of these methods is the following. Together with the object hologram on which the object distortions and aberrations of the interferometer are written, a reference hologram is obtained containing only aberrations of the interferometer. Subsequent combined optical processing of the object and reference holograms yields an interferogram of the object that is free from aberrations.

Let us consider the possibility of compensation of aberrations of the interferometer in the proposed method of interference flaw detection in periodic objects through the use of a reference hologram. We assume that the reference hologram is recorded in the plane *11* with interference of the reference wave (5) and object wave propagating along the optical axis of the object arm with the transparency *6* removed,

$$A'(x,y)=a_0\exp[i\psi(x,y)].$$

When recording the hologram under linear conditions, the transmittance of the hologram after chemical processing is given by

$$\tau_0(x',y')=1+\cos[2\pi n(\xi_k x'+\eta_k y')+\psi(x',y')]. \tag{13}$$

We assume that the reference hologram (13) is mounted in exactly the same place as before in the recording plane *11* and is illuminated by the object wave (11) and the plane wave $A''(x',y')=a_0$. Two waves propagate in the space behind the hologram along the normal to it:

$$B_1=b_1,\quad B_2=b_2\exp[in\varphi_k(x',y')],$$

where b_1 and b_2 are the amplitudes of the waves.

When these waves are filtered, they create an interference pattern

$$I(x'',y'')\sim 1+\cos[in\varphi_k(x'',y'')], \tag{14}$$

where $x''y''$ is the coordinate system chosen in the observation plane of the interferogram.

As can be seen from expression (14), aberrations are absent in the formation of the interference pattern obtained using a reference hologram (13). The The coefficient of increase of the measurement sensitivity, as in Eq. (6), is equal to the number of the selected diffraction order n.

Let us now consider how to obtain an interferogram of a periodic object with compensation of aberrations using the method of optically conjugate holograms without a reference beam with a reference hologram.[11] We assume that, as before, the reference hologram (13) is mounted in the recording plane *11*, where it is optically conjugate with the object *6*.

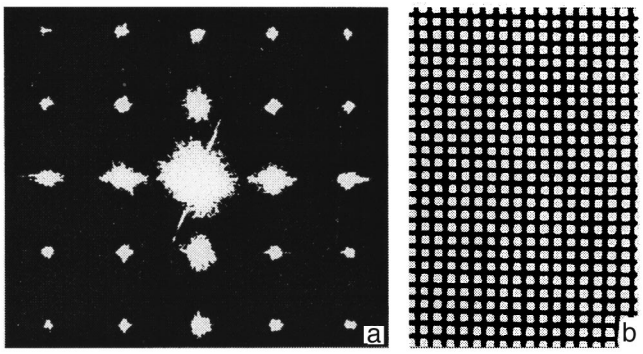

FIG. 2. Image of the diffraction spectrum of the screen (a) and a magnified image of a segment of the screen (b).

We also assume that waves are selected in the filtering plane *17* that have been diffracted by the object to the complex-conjugate orders

$$A_{kn}=a_{kn}\exp\left\{i\left[2\pi n\left(\frac{\cos\alpha_k}{\lambda}x+\frac{\cos\beta_k}{\lambda}y\right)+\psi(x,y)+n\varphi_k(x,y)\right]\right\}, \tag{15}$$

$$A_{kn}^{*}\sim a_{kn}\exp\left\{i\left[2\pi n\left(\frac{\cos\alpha_k}{\lambda}x+\frac{\cos\beta_k}{\lambda}y\right)+\psi(x,y)+n\varphi_k(x,y)\right]\right\}. \tag{16}$$

When the reference hologram (13) is illuminated by the waves (15) and (16), waves propagate along the normal to the hologram diffracted to the ± 1 orders, having the form

$$B_1'=\frac{1}{2}a_{kn}\exp[in\varphi_k(x',y')], \tag{17}$$

$$B_2'=\frac{1}{2}a_{kn}\exp[-in\varphi(x',y')]. \tag{18}$$

All the remaining waves formed in this process differ from waves (17) and (18) in their direction of propagation and can be removed from the beam with the help of a filtering diaphragm. Waves (17) and (18) form an interference pattern

$$I\sim 1+\cos[2n\varphi_k(x'',y'')]. \tag{19}$$

As can be seen from expression (19), this method can be used to compensate aberrations of the interferometer. In this case the coefficient of increase of the measurement sensitivity has grown by a factor of two in comparison with the interferogram (14).

EXPERIMENTAL VALIDATION

The above-described technique was validated using a metal screen consisting of two overlapping wire grids. The period of the grids was $T_1=T_2\approx 1$ mm, and the diameter of the screen was 125 mm. To obtain an image of the screen, its spatial spectrum, and interference patterns in various selected orders of the spectrum, the screen is mounted in position *6*.

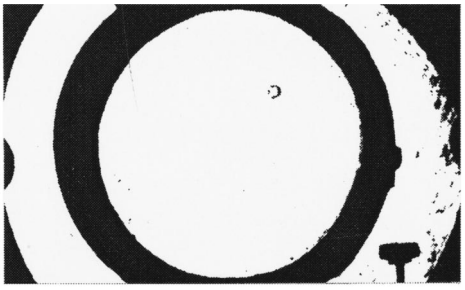

FIG. 3. Interferogram visualizing aberrations of the interferometer, obtained by isolating the zeroth-order wave.

Figure 2a displays the diffraction spectrum of the screen, recorded in the plane *17*. The working orders of the spectrum, corresponding to single diffraction on each grid, are arrayed along the vertical line and horizontal line intersecting at the point of the central, brightest beam—the zeroth-order beam. The remaining orders of the spectrum are a consequence of cross diffraction. Figure 2b displays a magnified image of a segment of the screen, recorded in the plane *11* with the diaphragm *17* removed. Since the two grids making up the screen overlap, the wave diffracted by either of these two grids to any diffraction order $n\neq 0$ contains information about the shape of the screen surface and defects in that surface. The wave corresponding to zeroth order does not carry phase distortions associated with the screen and can be used to estimate the magnitude of the aberrations of the interferometer. Figure 3 displays the illuminance distribution of the interference pattern recorded in the plane *11* tuned to an infinitely wide fringe to select the zeroth order at the diaphragm *17*. The magnitude of the aberration of the interferometer is 0.2λ. Figure 4a shows the illuminance distribution of the interference pattern for selection of the first-order wave diffracted by the vertical lines of the grid. Figure 4b shows the interference fringes obtained when selecting the second-order wave diffracted by the vertical lines of the grid. The increase in sensitivity is equal to two. As can be seen

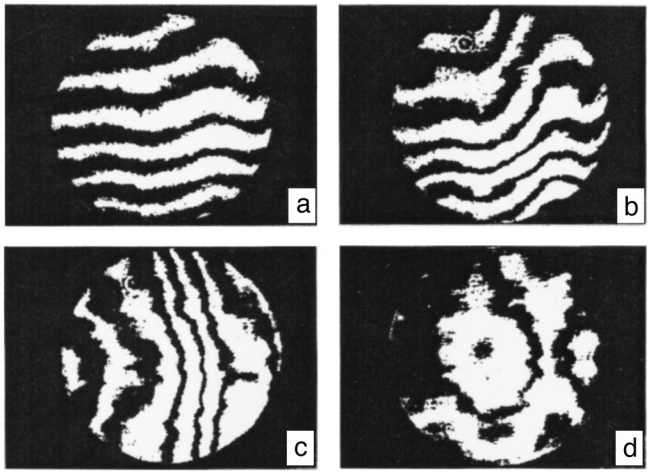

FIG. 4. Interferograms visualizing a defect in a metal screen by isolating the zeroth-order wave (a) and by isolating the second-order wave tuned to the horizontal fringes (b), the vertical fringes (c), and the fringe of infinite width (d).

from Fig. 4b, the increased displacement in this case of the tuning fringes in the defect zone permits a more accurate measurement of the magnitude and shape of the surface defect. Figures 4c and 4d show interference patterns obtained by selecting the second-order waves tuned, respectively, to the vertical fringes and the fringe of infinite width.

To summarize, the technique developed here for visualizing defects of periodic transmission objects allows one to increase the sensitivity and accuracy of measurement when recording interferograms of such objects in real time.

This work was supported by the Ministry of Public Education of the Republic of Belarus.

[1] I. Havanesian and Y. Y. Hung, Appl. Opt. **10**, 2734 (1971).
[2] N. Wadsworth, M. Marchant, and B. Billing, Opt. Laser Technol. **5**, No. 3, 119 (1973).
[3] A. M. Lyalikov, Opt. Zh., No. 1, 28 (1995) [J. Opt. Technol. **62**, 26 (1995)].
[4] Robert K. Erf (Ed.), *Holographic Non-Destructive Testing* [Academic Press, New York, 1975; Mashinostroenie, Moscow, 1979, 446 pp.].
[5] I. K. Gubarevich and A. M. Lyalikov, Opt. Spektrosk. **80**, 948 (1996) [Opt. Spectrosc. **80**, 855 (1996)].
[6] A. M. Lyalikov, Opt. Zh. **64**, No. 1, 82 (1997) [J. Opt. Technol. **64**, 68 (1997)].
[7] A. M Lyalikov, Opt. Zh. **63**, No. 5, 73 (1996) [J. Opt. Technol. **63**, 391 (1996)].
[8] N. M. Spornik, Opt. Mekh. Prom. , No. 2, 77 (1973).
[9] R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography* (Academic Press, New York, 1971).
[10] H. J. Caulfield (Ed.), *Handbook of Optical Holography* [Academic Press, New York, 1979; Mir, Moscow, 1982, Vol. 2, 735 pp.].
[11] A. K. Beketova, A. F. Belozerov, A. N. Berezkin *et al.*, *Holographic Interferometry of Phase Objects* [in Russian], Nauka, Leningrad, 1979.
[12] A. N. Zaĭdel' and G. V. Ostrovskaya, *Laser Methods of Plasma Research* [in Russian], Nauka, Leningrad, 1977, 232 pp.
[13] L. H. Tanner, Opt. Laser Technol. **4**, 281 (1972).

Translated by Paul F. Schippnick

BRIEF COMMUNICATIONS

Ball lightning with a lifetime $t \leqslant 1$ s

A. M. Boĭchenko

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia
(Submitted May 26, 1998; resubmitted October 6, 1998)
Zh. Tekh. Fiz. **69**, 131–134 (October 1999)

Among luminous formations acknowledged as examples of artificial ball lightning with lifetimes $t \leqslant 1$ s, examples are presented whose nature is well described by a weakly ionized plasma with a gas temperature $T \approx 0.5$ eV. It is shown that such lifetimes do not contradict the estimates of P. L. Kapitsa if it is taken into account the fact that the plasma under consideration is not an ideal blackbody. © *1999 American Institute of Physics.* [S1063-7842(99)01910-8]

INTRODUCTION

The term ''ball lightning'' is not clearly defined. In the broad sense, ball lightning is understood to mean an induced luminous formation observed under natural or artificial conditions. Studies of all possible luminous formations have lead to the conclusion that the nature of ball lightning is many-sided: ''...the impression is created that natural ball lightning is possibly not one, but many phenomena, similar in their manifestations, but with a different physical nature, different stability criteria, and somewhat different properties depending on the state of the atmosphere and the environment at the moment of observation of the event.'' (Ref. 1, p. 148).

In a previous paper[2] I advanced a hypothesis about the nature of bead lightning. According to what I wrote there, bead lightning is a weakly ionized plasma with a gas temperature of $T \approx 0.5$. The lifetime of the beads is governed by thermal conduction and correlates well with the observed dimensions of the beads ($t \approx 1$ s for a bead radius $R \approx 20$ cm). The maximum lifetime of such formations is of the order of a second, and such phenomena are fundamentally incapable of explaining ball lightning with a lifetime of the order of a minute, which is of the greatest interest.[3–5] But they are also interesting in regard to the nature of ball lightning since ''according to the available reports, ball lightning ''lives'' most often for 1–2 s. Lifetimes of such duration or less are noted in 80% of investigated reports'' (Ref. 1, p. 46). But might such instances of ball lightning with lifetimes $t \leqslant 1$ s also include objects described in Ref. 2? Such objects should arise during a powerful injection of energy into a medium whose temperature can amount to several electronvolts. In this case, the medium can be assumed to be almost completely ionized. In such a case, according to the estimates of Ref. 6, the formations under discussion should not exist for more than 10 ms. This circumstance has so far been a very weighty argument against the idea that ball lightning can have a high temperature (see, e.g., Ref. 4).

In the present Report, I show that in a more correct description the arguments of Ref. 6 do not contradict the results of Ref. 2. Artificially induced luminous formations will be indicated, whose nature is adequately described by the hypothesis advanced in Ref. 2. I consider only examples of formations for which detailed information about the conditions of their creation is available. Such cases are very few, but such formations are of interest, first of all, because they have long been acknowledged as classical examples of ball lightning.

ESTIMATE OF THE LIFETIME OF THE FORMATIONS

Reference 2 presents a kinetic treatment of processes taking place in a medium as a result of a powerful injection of energy. The arising objects, according to Ref. 6, should not exist for more than 10 ms. On the other hand, it follows from Ref. 2 that their lifetime is bounded by ~ 1 s. On the face of it, this seems to be a clear contradiction. The estimate in Ref. 6 was obtained by extrapolation of the parameters of a nuclear explosion to cases of ball lightning. This extrapolation contains the implicit assumption that in this case ball lightning is a perfect blackbody. However, if we take into account the difference in the emissivity of a medium with $T > 0.5$ eV from the emissivity of a perfect blackbody, then the estimate of the lifetime coincides with the estimate given in Ref. 2. Indeed, let us estimate the lifetime of luminous formations for air, proceeding from energy losses as a result of thermal emission:

$$\frac{7}{2} NV \frac{dT}{dt} = -\xi \sigma T^4 S.$$

Here N is the concentration of air molecules, V and T are the volume and temperature of the excited region, σ is the Stefan–Boltzmann constant, S is the surface area of the excited region, ξ is the thermal emission coefficient of air $\xi(T=1\text{ eV}) = 7 \times 10^{-3}$, $\xi(T=0.5\text{ eV}) = 10^{-3}$ (Ref. 7, p. 791).

Fitting ξ by a power-law dependence $\xi = 10^{-2} a T^{\gamma}$, we obtain $a = 0.7$, $\gamma = 2.8$. Taking the initial temperature

$T_0 \approx 1$ eV, $R(T_0) \approx 20$ cm, and imposing the condition $p = \text{const} = 1$ atm, we obtain $N \approx 10^{18}/T\ \text{cm}^{-3}$, $R \approx 20T^{1/3}$ cm. Then

$$\frac{dT}{dt} \approx -187 \cdot T^{7.47},$$

where t is expressed in seconds, and T in eV.

For the time required to cool down to the temperature T we have

$$t = -\int_{T_0}^{T} \frac{dT}{187 \cdot T^{7.47}} \approx \frac{8.26 \times 10^{-4}}{T^{6.47}}$$

and for $T < 0.4$ eV the cooling time $t > 0.32$ s. We note at once that t is bounded from above by a time of the order of a second, governed by thermal conduction,[2]

$$t \approx \frac{\chi}{(R/2.4)^2},$$

since for $T \approx 0.5$ eV we have $\chi = \nu_T/\sigma_{at} N \approx 100\ \text{cm}^2$, where $\sigma_{at} \approx 10^{-15}\ \text{cm}^2$ is the characteristic cross section of the atomic interaction, $\nu_T \approx 10^5$ is the mean thermal velocity of the gas particles, and $N \approx 10^{18}\ \text{cm}^{-3}$.

The degree of ionization α of the emitting volume can be estimated using the Sach–Boltzmann formula

$$N_e N_i / N = (g_e g_i / g)(m_e T_e / 2\pi\hbar)^{3/2} \exp(-J/T),$$

where j is the ionization potential of the gas; N_e, g_e, m_e, and T_e are the concentration, statistical weight, mass, and temperature of the electrons; and N_i and g_i are the concentration and statistical weight of the ions.

For $J = 14$–15.6 eV (nitrogen) and T_e, $T = 0.5$ eV we obtain $N_e \approx 10^{13} - 6 \times 10^{13}\ \text{cm}^{-3}$, equivalently, $\alpha = 10^{-5} - 10^{-4}$.

SOME EXAMPLES OF ARTIFICIAL BALL LIGHTNING

Let us turn now to some interesting artificial luminous objects.

1. The formation of fireballs in some submarines has been described in detail in Refs. 8 and 9 (see also Ref. 1, pp. 62 and 174). When the circuit breaker of the dc circuit of a system of storage batteries is shorted out, a return-current relay automatically disconnects the closed contacts, which leads to the formation of an arc between them. Usually, the arc quickly extinguishes. But sometimes a detached green fireball has been observed, moving in an outward direction away from the contacts. The floating balls existed for roughly 1 s, and their observed diameter varied roughly from 10 to 15 cm. The green light is apparently due to emission from neutral copper atoms vaporized from the copper contacts during formation of the fireball. By control experiments it was established that the fireballs arose only under those conditions in which certain threshold values of the power or the current in the circuit were exceeded. For example, a small fireball arose for a current $I = 1.5 \times 10^5$ A and voltage $V = 260$ V. According to refined data, the energy contained in the fireball was $E = 400$ J (Ref. 1, p. 63). The described regularities were confirmed by tests of return-current devices at the Philadelphia Naval Shipyard in 1974.

If the hypothesis of Ref. 2 is valid, then the energy of the fireball contains translational, rotational, and vibrational degrees of freedom of the particles that make up the fireball. Then

$$E = i\hat{N}T,$$

where $\hat{N}$ is the total number of particles in the fireball (since fireballs appear in air, I assume that the particles are mainly diatomic molecules or atoms, and accordingly $i = 7/2 - 3/2$, depending on the degree of dissociation of the air molecules).

The described processes occur in the atmosphere; therefore, from the equality of pressures inside and outside the fireball we have

$$NT = N_0 T_0,$$

where N, N_0 and T, T_0 are the particle concentrations and temperatures inside and outside the ball.

For the above value of E the radius of the balls should be equal to

$$R = \sqrt[3]{\frac{3V}{4\pi}} = \sqrt[3]{\frac{3\hat{N}}{4\pi N}} = \sqrt[3]{\frac{3E}{4\pi i N_0 T_0}}$$

and should equal $R = 8.3 - 6.2$ cm for $i = 3/2 - 7/2$, which agrees well with the observations. The considered formations are presented in Refs. 1, 3, and 4 as examples of ''artificial'' ball lightning.

2. A large number of works have been dedicated to a description of plasmoids, obtained in high-frequency and microwave discharges (see, for example, Ref. 1, pp. 125–129). In the overwhelming majority of cases, the plasmoids disappear almost immediately after the electromagnetic field is switched off. Apparently, the first long-lived plasmoid at atmospheric pressure was obtained in Refs. 10–12 (see also Ref. 1, p. 182). A ball of diameter 15 cm existed for 0.5–1 s after termination of a high-frequency excitation. These experiments are also cited in Refs. 1 and 3 as examples of artificial ball lightning. Here, however, I consider the plasmoids described in Ref. 13, since the experiments cited there were conducted with very good diagnostic equipment. The diagnostics made it possible not only to measure directly some parameters, but also, with the help of numerical calculations, to obtain those parameters whose direct measurement is impossible.

The experiments were conducted in air on different setups. Variation of the setups made it possible to vary the energy flux density of the focused beam of electromagnetic radiation in the focal region S, the wavelength, the half-maximum duration of excitation τ, and the pulse repetition rate. Below I cite only those facts that are of interest in connection with the given work and in which excitation of the medium was realized with one pulse.

a) A study of the velocity of propagation of a discharge in air was carried out where the discharge was initiated by a spark with $S = 10^3\ \text{W/cm}^2$ and $\tau = 800\ \mu$s (Ref. 13, pp. 161–162). The atmospheric-pressure discharge was a uniformly luminous plasma formation during the entire time of its existence. The gas was heated to high temperatures of the order

of $5000-6000$ K (≈ 0.5 eV). The afterglow time of the discharge was 1 ms. The speed of propagation of the discharge during the first microseconds of its existence reached values of the order of 10^5 cm/s, decreased abruptly, and at the end of the pulse dropped to values $\approx 6\times 10^2$ cm/s. Taking the mean velocity of propagation v to be a few thousand cm/s, we find that the radius of the excited region should be $R\approx v\tau\approx 1$ cm.

b) For $S=10^4$ W/cm^2 and $\tau=800\,\mu$s the time dependence of the gas temperature of the plasma of an equilibrium microwave discharge at atmospheric pressure was measured (Ref. 13, pp. 171–172). Results of measurements obtained using an optical technique are in good agreement with values of the temperature temperature calculated by Sach's formula under the assumption of local thermodynamic equilibrium using a time dependence of the electron density that was measured experimentally under the same conditions. At the end of the pump pulse the gas temperature was equal to $T\approx 6000$ K. After termination of the microwave pulse, the gas slowly cools with a characteristic time ≈ 1 ms. The spatial distribution of the electrons N_e is close to table-shaped. If we estimate the radius of the excited region from the radius of the region in which N_e is nonzero, then its radius will be $R\approx 0.8$ cm.

Note that the thermal-conduction cooling time of the excited region $R\approx 1$ cm is of the order of 1 ms, the time obtained in cases (a) and (b). In the experiments, larger plasmoids were also obtained, with radii $R\approx 20-30$ cm.

c) In a study of the dynamics of the development, existence, and decay of a plasma in a discharge initiated with the help of a metal–insulator contact placed in the region of the focus of a microwave beam, with $S=10^3$ W/cm^2 and $\tau=100$ ms at atmospheric pressure, a more or less uniform luminous formation was obtained, elongated in the direction of the focusing antenna (Ref. 13, pp. 174–175). At the end of the pump pulse the dimensions of the plasma reached a value ≈ 60 cm in the longitudinal direction and ≈ 15 cm in the direction of the electric field vector **E** in the incident wave. After termination of the pulse, the region of space where the microwave discharge existed continues to glow brightly for ≈ 0.3 s. The gas temperature was $\approx 5000-7000$ K. The heated glowing volume of air floats upward (due to the buoyancy force) and, gradually cooling, floats out of the field of view of the television camera.

d) For pumping with $1<\tau<10$ s, a vertical feed of microwave energy to the discharge from below was used (Ref. 13, pp. 179–182). Motion of the discharge downward counter to the delivered energy was compensated by motion upward due to convective removal of heat. For $S=10^3$ W/cm^2 and $\tau=600$ ms intense convective flows of heated luminous air are formed above the discharge and ≈ 300 s after initiation of the excitation, the discharge takes on a mushroom-like shape. With time, the cap of the mushroom floats upward, detaches from its foot, and floats out of the field of view of the camera. After the microwave field is switched off, the plasma continues to glow brightly for hundreds of milliseconds. When S is lowered to 300 W/cm^2 it becomes possible to keep the discharge stationary for the entire duration of the pump pulse $1.5<\tau<10$ s. For $S=200$ W/cm^2 a regime of motion is observed in which the plasma ball has a diameter of $10-30$ cm and floats vertically upward with a velocity of ≈ 1 m/s.

Note that a floating-upwards, similar to that described for cases (2c) and (2d), was also observed for the fireballs in case (1) and for the plasmoids in Refs. 10–12, which also gives an idea of the high temperature inside these formations. The existence time of the formations for cases (1), (2c), and (2d) is $0.2<t<1$ s and coincides with the ermal-conduction cooling time for the excited regions with $R\approx 10-20$ cm mentioned in the Introduction.

CONCLUSION

Among the artificially created luminous formations considered in the literature as examples of ball lightning with lifetimes $t\leqslant 1$ s, I have singled out formations whose nature can be well explained if one assumes that they are a weakly ionized plasma with a gas temperature $T\approx 0.5$ eV (6000 K). Their maximum lifetime, estimated in Ref. 6, after a correction for the temperature dependence of the coefficient of thermal emission of air, is consistent with the lifetime obtained in Ref. 2, which is equal to $t\approx 0.2-1$ s for radii of the balls $R\approx 10-20$ cm.

[1] J. D. Barry, *Ball Lightning and Bead Lightning: Extreme Forms of Atmospheric Electricity* [Plenum Press, New York, 1980; Mir, Moscow, 1983].
[2] A. M. Boĭchenko, Fiz. Plazmy **22**, 1012 (1996) [Plasma Phys. Rep. **22**, 917 (1996)].
[3] S. Singer, *Nature of Ball Lightning* [Plenum Press, New York, 1971; Mir, Moscow, 1973].
[4] I. P. Stakhanov, *On the Physical Nature of Ball Lightning* [in Russian], Nauchnyĭ Mir, Moscow, 1996.
[5] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Usp. Fiz. Nauk **164**, 297 (1994).
[6] P. L. Kapitsa, Dokl. Akad. Nauk **101**, 245 (1955).
[7] I. S. Grigor'ev and E. Z. Meĭlikhov (Eds.), *Handbook of Physical Constants* [in Russian], Énergoatomizdat, Moscow, 1991.
[8] P. A. Silberg, J. Geophys. Res. **67**, 4941 (1962).
[9] P. A. Silberg, in *Problems of Atmospheric and Space Electricity*, edited by S. C. Coroniti (Elsevier, Amsterdam, 1965), p. 336.
[10] J. R. Powell, D. Finkelstein, M. S. Zucker, and J. R. Manwaring, ''Laboratory production of self-sustained atmospheric luminosities,'' Eighth Annual Meeting, American Physical Society, 1966.
[11] J. R. Powell and D. Finkelstein, in *Advances in Geophysics* (Academic Press, New York, 1969), Vol. 13, p. 141.
[12] J. R. Powell and D. Finkelstein, Ann. Rheum. Dis. **58**, 262 (1970).
[13] A. S. Zarin, A. A. Kuzovnikov, and V. M. Shibkov, *A Freely Localized Microwave Discharge in Air* [in Russian], Neft' i Gaz, Moscow, 1996.

Translated by Paul F. Schippnick

Transport of metals in Si by the thermal effect of continuous CO_2 laser radiation

P. S. Shkumbatyuk

Drogobych State Pedagogic Institute, 293720 Drogobych, Ukraine
(Submitted April 2, 1998)
Zh. Tekh. Fiz. **69**, 135–137 (October 1999)

The influence of the local thermal effect of continuous CO_2 laser radiation with a power density as high as 10^3 W/cm^2 on metal–Si films is investigated. It is discovered that the structural defects formed as a result of irradiation influence the transport of the metal in Si. © *1999 American Institute of Physics.* [S1063-7842(99)02010-3]

Among the various technological approaches for implanting impurities in Si, utilization of the thermal effect of optical radiation for doping Si with metals by irradiation of a metal–Si film has been studied most thoroughly. When pulsed radiation is employed, liquid-phase implantation is promoted by a thermal mechanism, and the solid-phase implantation of metal atoms can be influenced by the pressure of laser-induced shock waves.[1] In addition, it should be expected that investigations of the local thermal effect of continuous optical radiation can provide additional information regarding the mechanism of the implantation of metals in Si.

With this aim we report here the results of an investigation of the implantation of a metal in Si using continuous CO_2 laser radiation with a power density up to 10^3 W/cm^2 acting on the metal–silicon system. The substrate used in this study was p-Si with $p = 1 \times 10^2 \Omega \cdot$ cm and a thickness of 0.6–0.8 mm, and the metal (Al or Au) was deposited on the silicon surface in the form of particles with a total mass of up to 2 mg. The irradiation was carried out at room temperature in air. No melting of the original material was detected at the power density indicated.

Figure 1 shows cleavage surfaces at different depths after exposure of an Al–Si film to laser radiation with a power density of 8×10^2 W/cm^2. The mass of the metal exceeded 1 mg. The cleavage surfaces in the laser-irradiated zone were polycrystalline, and no macroscopic inclusions of the metal were detected across the thickness of the original sample. It was found that the metal emerged on the surface opposite to the irradiated surface, and for the ''through'' layer in the irradiation zone the conductance increased with decreasing temperature. Structural defects in the form of microcracks are revealed at a depth of 0.1 mm from the surface in the irradiation zone (Fig. 1a), and their concentration is lower at a depth of 0.3 mm (Fig. 1b). A systematic analysis of the irradiation of the Al–Si system revealed that the formation of a polycrystalline structure with emergence of the metal on the opposite surface has several stages: formation of a metal drop with an oxidized surface, interaction of the metal with the surface in the irradiation zone, and the appearance of Si particles in the metal. When less than 0.5 mg of the metal was deposited, its emergence on the surface opposite to the irradiated surface was not detected, regardless of the parameters and time of action of the laser radiation. When the sample was irradiated on the side opposite to the metal–Si surface, emergence of the metal on the irradiated surface did not depend on the amount of metal deposited.

An analysis of the effect of laser radiation on the Au–Si system for various starting conditions and a comparison with the Al–Si system without allowance for the exposure time reveal that no melting of the metal occurs in the subsurface region of Si. Microchannels are detected on a cleavage surface across the thickness of the irradiation zone (Fig. 2), and Au inclusions may form in the channels, depending on the amount of metal.

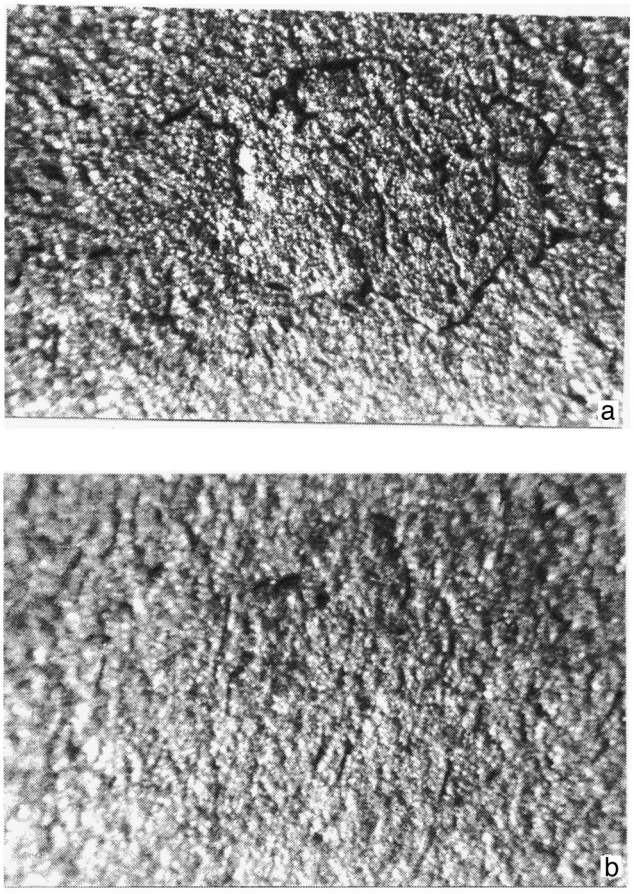

FIG. 1. Structure of an irradiated Al–Si sample at depths of 0.1 (a) and 0.3 mm (b) from the irradiated surface.

Let us analyze a possible model of the implantation of a metal in Si on the basis of these data. Melting of the metal on the Si surface can occur when Si is heated, if the cross-sectional area of the laser beam is greater than that of a metal particle or the absorption of radiation by the metal increases as a result of its oxidation.[2] The presence of particles of the material in the metal indicates that they interact in the surface region in the case of metals which form silicides. The formation of defects with an increased concentration in the subsurface region of a metal–Si film can be caused by the influence of thermal stresses with consideration of the additional quantity of heat evolved in a local region of Si by the metal:

$$Q(T) - cm\alpha L(T - T_0), \tag{1}$$

where c and m are the specific heat and mass of the metal, α is the thermal conductivity of Si, L is the melting depth, and T_0 is the temperature outside the irradiation zone.

An estimate of the thermal stresses with consideration of the Gaussian distribution of the temperature in the irradiation zone $T_{(r)} = T\exp(-r^2/a_0^2)$ can be determined after Ref. 3. In the case of axial symmetry, the diagonal components of the stress tensor are determined from the equations

$$\Sigma_{rr} = \frac{T\vartheta E}{1-\nu}\frac{a^2}{r^2}\left[\exp\left(-\frac{r^2}{a^2}\right) - 1\right], \tag{2}$$

$$\Sigma_{\varphi\varphi} = \frac{T\vartheta E}{2(1-\nu)}\frac{a^2}{r^2}\left[1 - \left(1 + \exp\frac{2r^2}{a^2}\right)\exp\left(-\frac{r^2}{a^2}\right)\right], \tag{3}$$

where T is the temperature in the laser-irradiated zone, ϑ is the coefficient of linear expansion, ν is Poisson's ratio, a_0 is the cross-sectional radius of the beam, and E is Young's modulus.

For an Al–Si film at $T = 1200\,^\circ$C (determined by a pyrometer) and $a_0 = 1.25$ mm, we obtain $\Sigma_{rr} = -8.17 \times 10^2$ kg/cm^2 and $\Sigma_{\varphi\varphi} = -7.39 \times 10^3$ kg/cm^2. According to Ref. 4, the theoretical value of $\Sigma_{\varphi\varphi}$ is greater than the limiting value of the compressive strength of Si, indicating the presence of thermal stresses, which lead to the formation of microcracks in the subsurface region. If it is taken into account that the formation of microcracks in crystalline materials is promoted by the relaxation of dislocation clusters, it is likely that the concentration of dislocations should be greater in the subsurface region and decrease across the thickness when an Al–Si film is heated by radiation, in agreement with Figs. 1a and b. Therefore, the defects formed during irradiation have a dominant influence on the transport of the metal. In the Au–Si system the presence of defects in the form of microchannels is possibly caused by the constraints on the formation of silicides and the solubility of Au. Nevertheless, with consideration of (1), when the amount of metal is small and the sample is irradiated from the metal–Si side, the possibility of the emergence of the metal on the opposite surface is probably restricted by the defect concentration gradient between the subsurface region and the bulk.

To analyze the lack of a dependence of the emergence of the metal on its amount with consideration of an arbitrary radiation absorption depth as a function of the Si temperature in the case of irradiation from the opposite side, we considered an additional mechanism, which is associated with the possibility of an accelerating influence of temperature fluctuations on the diffusion of the metal along the direction of the laser beam. For this purpose we investigated the transmission of the laser radiation through Si. According to Ref.

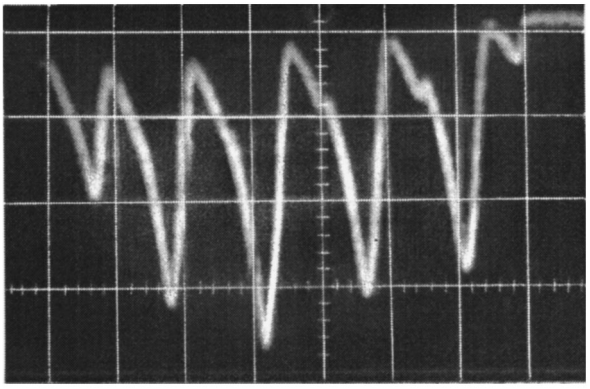

FIG. 3. Dependence of the transmission of radiation with a power density of 40 W/cm^2 in Si ($x = 2$ s per scale division).

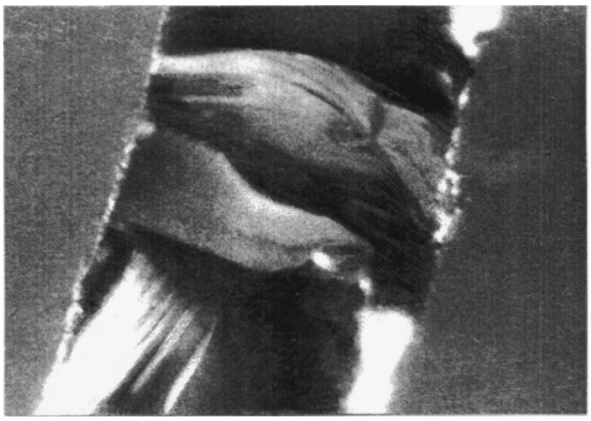

FIG. 2. Structure of an irradiated Au–Si sample across its thickness.

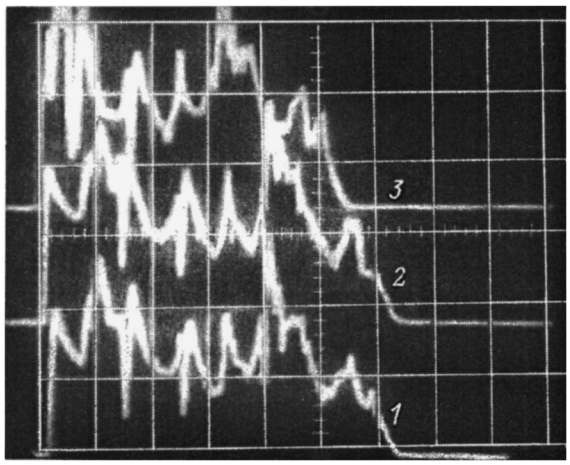

FIG. 4. Dependence of the transmission of radiation with a power density of 8×10^2 W/cm^2: *1,2*—$T_1 = T_2$; *3*—$T_3 > T_2$; T is the temperature of the Si before irradiation ($x = 0.2$ s).

5, the transmission coefficient is a function of the total number of charge carriers

$$T=\left[1+\frac{(\varepsilon_0-1)^2}{4\varepsilon_0}\sin^2\phi\right]^{-1}, \quad \phi=\phi_0+\Delta\phi_{(t)}, \tag{4}$$

$$\phi_0=\frac{\omega\sqrt{\varepsilon_0}}{c}h\left(1-\frac{2\pi e^2 n_0}{\varepsilon_0 m_n\omega^2}\right),$$

$$\Delta\phi(t)=-\frac{2\pi e^2}{\omega c\sqrt{\varepsilon_0}}\left(\frac{1}{m_n}+\frac{1}{m_p}\right)P_t(T) \tag{5}$$

[where h is the thickness of the plate, m_n and m_p are the electron and hole effective masses, and $P_t(T)$ is the time-dependent total number of electron-hole pairs per unit area of irradiated surface] and has a harmonic form when $P_t(T)$ is an exponential function. An investigation of the transmission of radiation with a power density equal to $40\,\mathrm{W/cm^2}$ with consideration of $P_t(T)=\sigma_t(T)$, where $\sigma_t(T)=\sigma_0\times\exp(-E_g/2kT)$, showed that the transmission coefficient agrees with Ref. 5 (Fig. 3). When the irradiation is carried with a power density of $8\times10^2\,\mathrm{W/cm^2}$, the transmission coefficient exhibits anharmonic behavior (Fig. 4), which indicates the nonexponential form of $P_t(T)$. It is theorized that the cause of such a form of $P_t(T)$ is the variation of the temperature in the irradiation zone. As a result, there are increased possibilities for the formation of defects in the bulk and a decrease in the defect concentration gradient between the irradiated and unirradiated surfaces with an increase in the irradiation zone; this leads to transport of the metal along the defects.

[1] N. N. Rykalin, A. A. Uglov, and N. V. Zuev, *Laser and Electron Beam Material Processing: Handbook* (Nauka, Moscow (1985), 250 pp.).
[2] A. M. Bonch-Bruevich, V. G. Dorofeev, and M. N. Libenson *et al.*, Zh. Tekh. Fiz. **52**, 1133 (1982) [Sov. Phys. Tech. Phys. **27**, 686 (1982)].
[3] B. A. Boley and J. H. Weiner, *Theory of Thermal Stresses* [Wiley, New York (1960) Mir, Moscow (1964)].
[4] P. I. Baranskiĭ, V. P. Klochkov, and I. V. Potykevich, *Semiconductor Electronics* [in Russian], (1975), 704 pp.
[5] A. I. Liptuga, V. K. Malyutenko, and I. I. Boĭko, Fiz. Tekh. Poluprovodn. **12**, 1447 (1978) [Sov. Phys. Semicond. **12**, 856 (1978)].

Translated by P. Shelnitz

Quantitative reference-free express analysis of some alloys on a laser time-of-flight mass spectrometer

G. G. Managadze

Space Research Institute, Russian Academy of Sciences, 117810 Moscow, Russia

N. G. Managadze

Institute for Problems of Control, Russian Academy of Sciences, 117806 Moscow, Russia

(Submitted February 9, 1998; resubmitted April 1, 1999)
Zh. Tekh. Fiz. **69**, 138–142 (October 1999)

A technique is proposed for analyzing the isotopic and elemental content of metals and alloys with the help of the LAZMA time-of-flight laser mass spectrometer. It is shown that for multicomponent alloys when working in the regime of the ''main'' maxima and recording all the ions in the energy spectrum the device enables one to perform quantitative measurements of sample composition and obtain valid results without the use of reference samples. © *1999 American Institute of Physics.* [S1063-7842(99)02110-8]

The compact laser time-of-flight mass reflectron LAZMA, developed at the Space Research Institute of the Russian Academy of Sciences[1–4] operates in the ion free path regime.[5] The instrument combines the high analytical, technical, and operational characteristics necessary to investigate the elemental and isotopic composition of solid-state samples, including powder samples by providing a mass resolution of 300–600, a sensitivity of roughly 1–10 ppm, and high reproducibility of the spectra for a homogeneous sample.

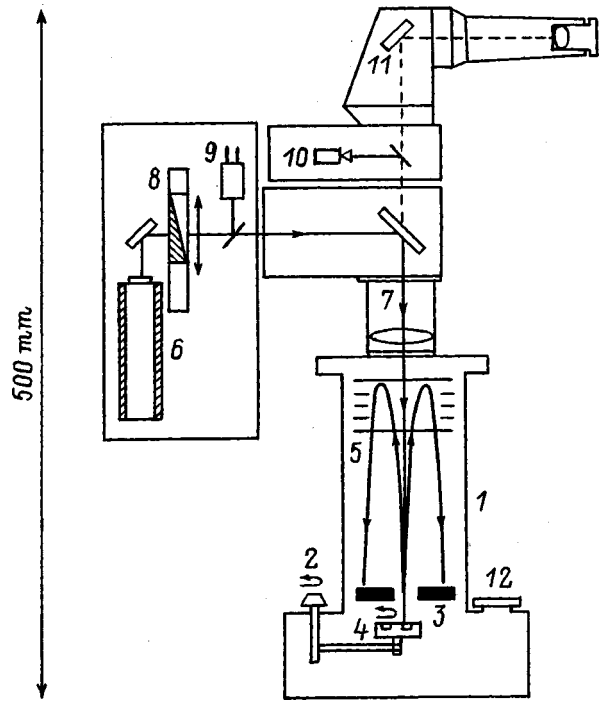

FIG. 1. Diagram of the LAZMA device: *1*—vacuum chamber, *2*—system for interchanging samples, *3*—detector, *4*—carriage with target, *5*—mass reflectron, *6*—laser, *7*—focusing lens, *8*—neutral filter, *9*—radiant power meter, *10*—target illuminator, *11*—microscope, *12*—air-lock chamber for introducing the sample.

The design of the analyzer device depicted in Fig. 1 consists of a mass-reflectron with a completely linear configuration possessing axial symmetry relative to the laser beam and motion of the formed ions. Laser light falls upon the surface of the sample, passes through the reflector grid and the opening in the detector. Such a design provides not only high reproducibility of the spectra, but also the possibility of carrying out a layer-by-layer analysis practically without limits. Variation of the power density of the laser radiation is effected with the help of a neutral-density filter and allows one to obtain singly ionized ions.

The developed technique of performing quantitative measurements[6] is based on the physical principle, according to which valid results are obtained by recording all primarily singly-charged ions formed by the laser beam in the energy interval from 0 to $E_{\max}$ (Ref. 5). Indeed, the laser ion source[6] provides not only nonfractional evaporation of the sample, but also an equiprobable yield of ions of different elements.

In the case under consideration, for the diameter of the laser beam equal to 50–60 μm and power density equal to

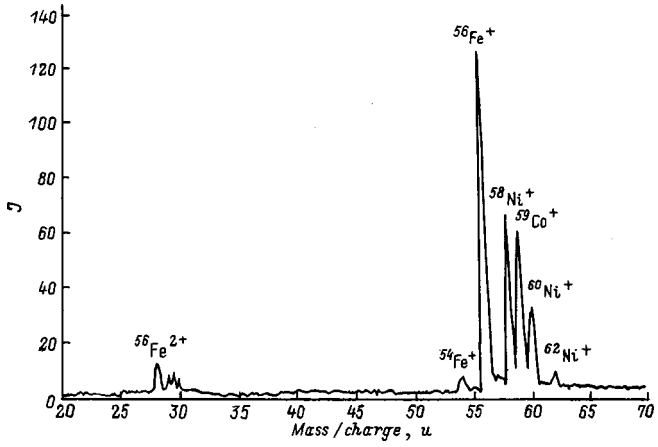

FIG. 2. Spectrum of Covar.

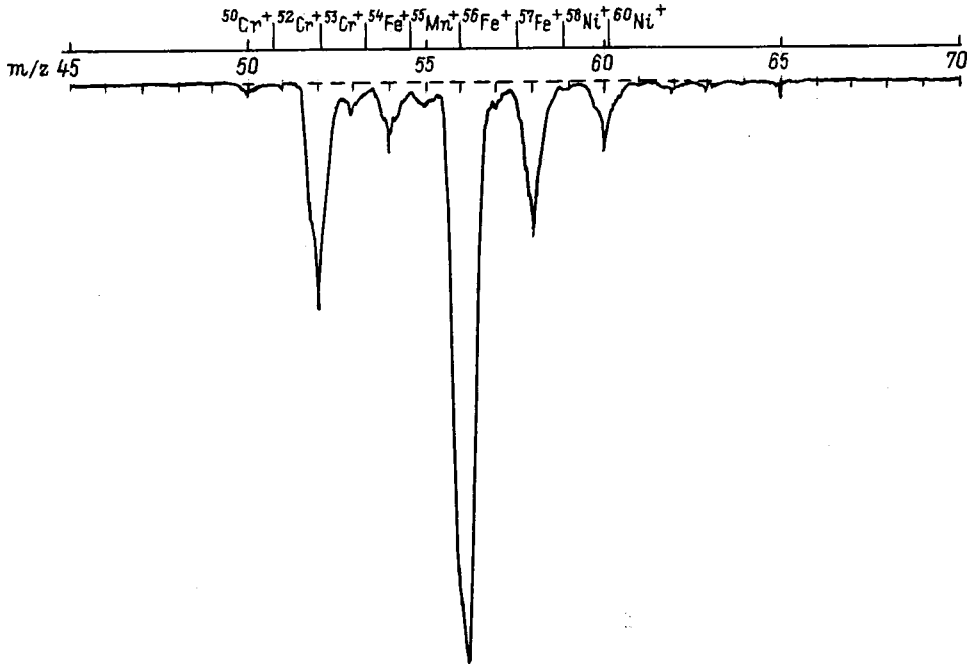

FIG. 3. Spectrum of stainless steel.

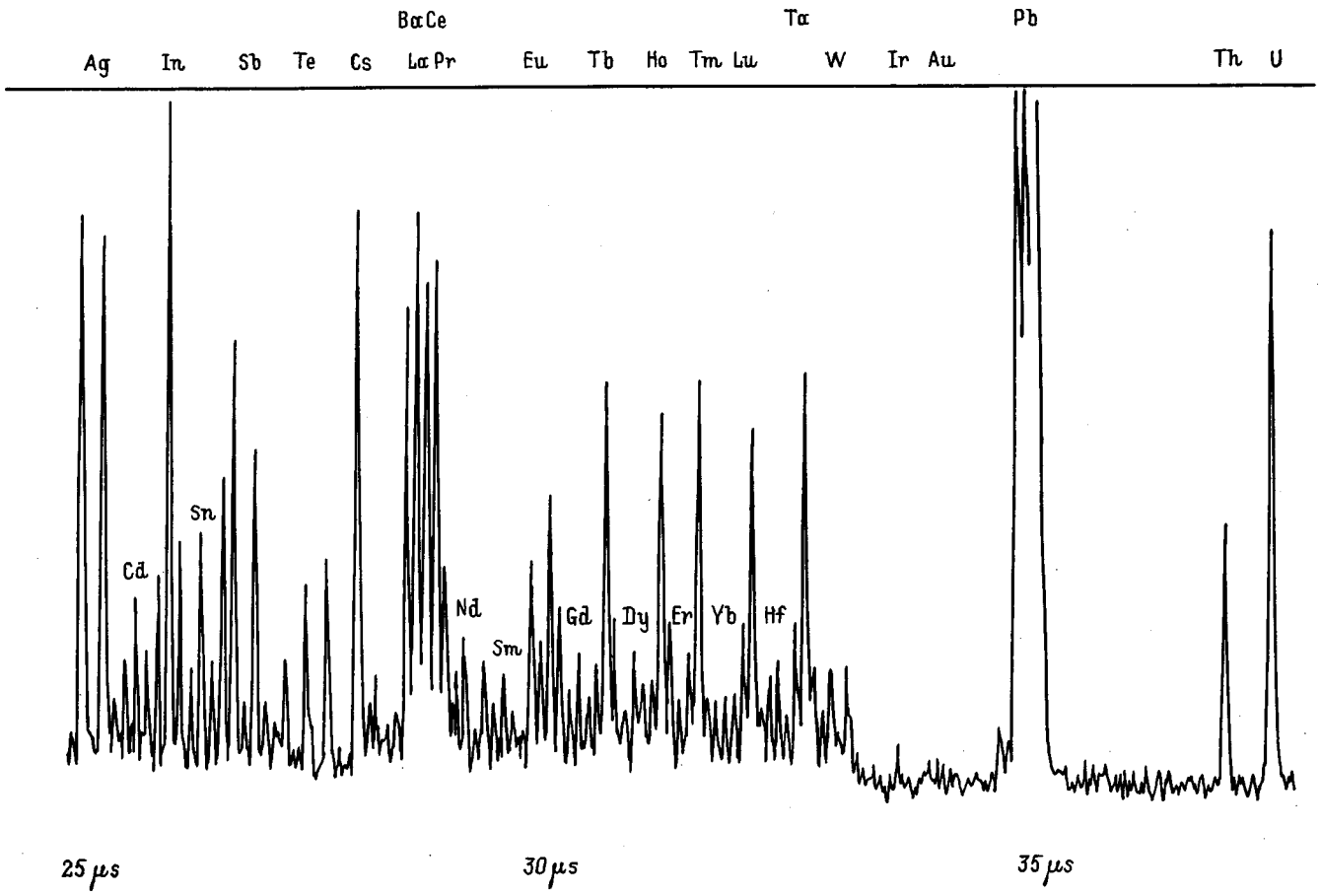

FIG. 4. Fragment of the spectrum of the standard glass NBS-612.

TABLE I. Elemental content of the Covar target in atomic %).

Element	Content according to the ''Metals Handbook''	Measured content in sample 1	Measured content in sample 2
Fe	52.5–54.5	53	53
Ni	28.5–29.5	30.7	30.5
Co	17–18	16.4	16.5

10^9 W/cm^2, ions of the majority of elements are concentrated in the energy interval from 0 to 150 eV. Consequently, when recording ions in this energy window the results should be valid, and the ratio of the measured value of the concentration of the element to its real value should be close to or equal to unity.

As measurements performed on various mass spectrometers have shown,[5] the above statements are valid for a wide class of devices with laser ion sources. The results of the present work also confirm this. However, when using this method it turned out that expanding the energy window leads to a sharp drop in the overall mass resolution.

Consequently, over a wide energy window fairly accurate results are obtained for the quantitative amount of various elements, but because of the low mass resolution the isotopes, and sometimes also the mass peaks of individual elements, coalesce. In the nominal regime of the device in the narrow energy window from 10 to 40 eV and with high mass resolution, accurate results are obtained only for elements of similar mass while for elements that are widely separated in mass the results differ substantially.

Given the above constraints, the optimal procedure here would be combined processing of data in the wide- and narrow-window regimes. In the wide-window regime the total amount of all the formed ions with energies from 0 to $E_{\max}$ is determined, and this amount is set equal to 100%. From these data the percent distribution of elements is determined from the individual peaks or from groups of peaks if the peaks are not resolved. Measurements are then performed in the narrow-window regime with high resolution and the relative quantities are determined for the previously unresolved elements or neighboring peaks using percent-content data on individual peaks or groups, obtained in the wide-window regime.

Combined processing of the results makes it possible to determine the true content of elements in a matrix, and also the content of impurities, taking into account the fact that neighboring elements have an equiprobable yield.

This is the gist of the proposed technique of reference-free analysis of alloys, which, without a doubt, requires experimental confirmation.

To investigate the analytical characteristics of the device in the narrow window, we used samples possessing high homogeneity with elemental concentrations ranging from 50% down to 0.05% (500 ppm): Covar, a standard steel, and the standard glass NBS-612, which contained homogeneously distributed metals in the form of impurities.

Figures 2–4 present spectra of Covar (Table I), steel (Table II), and a fragment of the spectrum of the standard glass NBS-612 for metals in the range from rubidium to rhenium.

The steel spectrum gives a picture of the yields of neighboring elements in the narrow-window regime. It can be seen that all elements are well resolved. The peak-integrated data on the quantitative content of elements in the matrix of the steel alloy turned out to be close to the standard values given in the literature.

Let us consider the results of a quantitative analysis of a multicomponent alloy in the wide-window regime. Toward this end, we selected a sample of industrial 14-karat gold an alloy of gold, silver, copper, and zinc). Besides the possibility of obtaining valid results in the wide-window regime here we also pursued the goal of recording the instrument function of the device for various elements widely separated in mass. In Fig. 5 two diverging straight lines bound the region of the actual content of metals in 14-karat gold, and the symbols represent the experimentally obtained values, the x's demarcate the root-mean-square errors of the results. In these graphs the summed zinc and copper content is plotted.

It is clear from these data that in the wide window the measured values of the content of the various elements in 14-karat gold are in good agreement with their actual values taken from the alloy certificate. This confirms the validity of the assumption that in the wide window the yields of different metals, including those widely separated in mass. This makes it possible to perform quantitative measurements with the LAZMA device.

To obtain complete information about the sample in the narrow window, we assayed the copper and zinc separately. These results are plotted in Fig. 6. The relative amounts of these neighboring elements was determined using the well-resolved mass spectra obtained in the narrow-window regime, and the content of each element in particular was determined using the previously obtained values of the percent content of the two metals taken together.

As can be seen from Fig. 6, the results of the measurements stand in good agreement with the standard data. It follows from the experimental data that the data of the first of the four series of measurements, the regime 10-147-93/94B approach the standard values the most closely.

Thus, the main goal of this work was achieved, namely,

TABLE II. Elemental content in a steel sample based on nominal and experimental data.

	Elemental composition in atomic %)							
Data	C	Mn	P	S	Si	Cu	Ni	Cr
Nominal C1154 19Cr-13Ni standard)	0.086	1.42	0.06	0.053	0.50	0.40	12.92	19.06
Experimental using the LAZMA device)	+	1.70	+	+	+	0.39	12.70	19.50

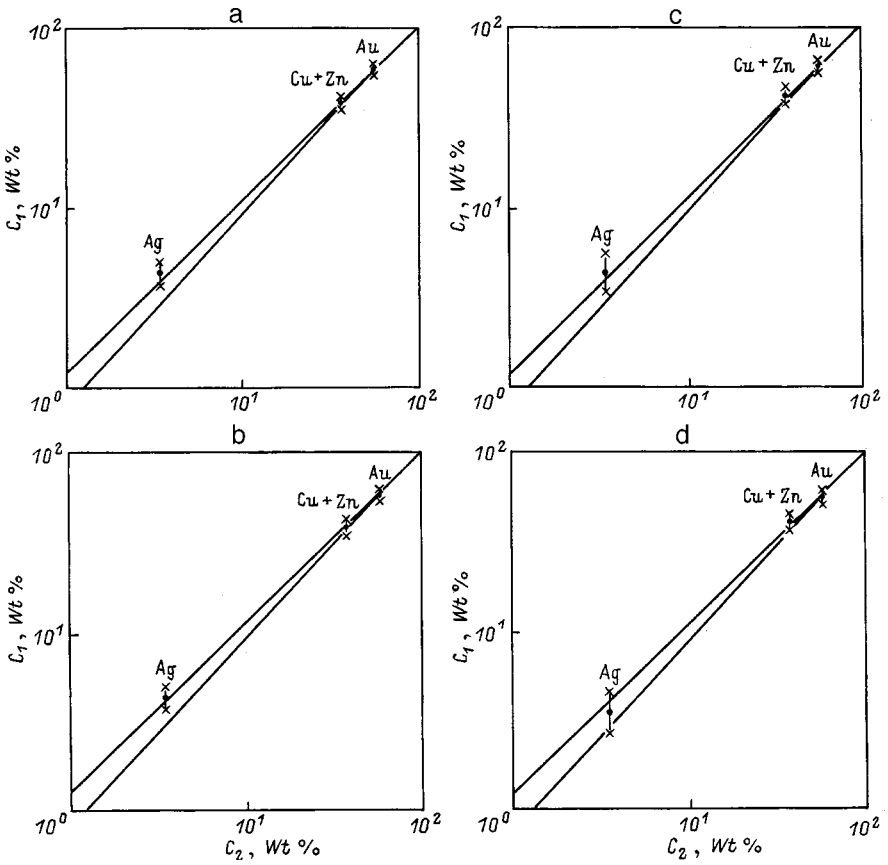

FIG. 5. Relation between elemental concentration levels in a gold alloy based on experimental data (C_1) and according to specifications (C_2), obtained for different voltages on U_2: a) 93–94, b) 95, c) 96, and d) 97–98 V. The voltage on $U_R = 147$ V. The data on copper and zinc are plotted as their total sum.

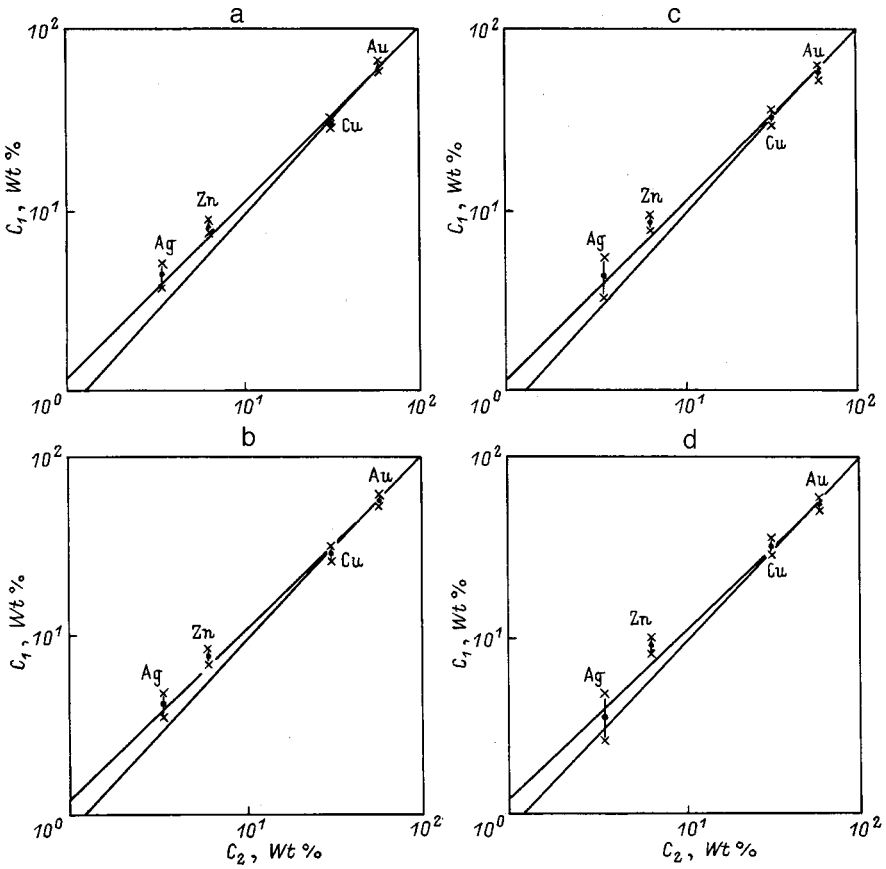

FIG. 6. Same as in Fig. 5, but with the data for copper and zinc plotted separately.

to demonstrate the possibilities of performing a chemical analysis of substances in the first phase, alloys) without the use of a reference sample and without sample preparation. It is not so important that the measurements were performed on a relatively convenient target—a homogeneous gold alloy; it is more important that the operativity of the technique was confirmed and that after processing spectra obtained in both the wide and narrow windows it is possible to correctly determine the composition of a multicomponent alloy.

We have also empirically confirmed the absence of a dependence of the efficiency of recording of the ions on their masses. From the obtained data it is hard to judge whether this characterizes only the given device or, possibly, this is a more general characteristic of a detector that receives ions accelerated to an energy of 2 keV before the first plate.

A continuation of this work is planned in which it is envisaged to examine nonmetallic samples. Of especial interest are multicomponent samples: ceramics, glasses, various minerals, and soils. After the completion of these measurements it will be possible to more accurately characterize the real analytical possibilities of the LAZMA device.

The authors express their gratitude to L. E. Chumikov for assistance provided during the completion of this work.

[1] G. G. Managadze, Inventor's Certificate USSR) No. 1732396 ; pub. Byull. Izobret., No. 17 , (1992).
[2] G. G. Managadze and I. Yu. Shutyaev, Chem. Anal. **124**, 505 (1993).
[3] R. McEntire, G. Managadze, and H. Rosenbauer *et al.*, *Forty-Sixth International Astronautical Congress*, Oslo (1995).
[4] G. G. Managadze, Preprint [in English], Space Research Institute, Russian Academy of Sciences, Moscow (1994).
[5] Yu. A. Bykovskiĭ and V. N. Nevolin, *Laser Mass Spectrometry* [in Russian], Énergoatomizdat, Moscow (1985), 129 pp.
[6] G. G. Managadze and N. G. Managadze, IKI RAN Preprint [in Russian], Space Research Institute, Russian Academy of Sciences, Moscow (1997).

Translated by Paul F. Schippnick