TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Vector solitons in the dynamics of anharmonic monatomic lattices

V. V. Bryksin

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

(Submitted October 9, 1996)
Zh. Tekh. Fiz. **68**, 1–6 (November 1998)

It is shown that three types of solitary acoustic waves can develop in anharmonic crystal lattices corresponding to the three branches of acoustic phonons. A system of three nonlinear Schrödinger equations is derived to describe this situation. For greatly different group velocities, the interaction between solitons reduces collisions between them. When the group velocities of the different acoustic modes in a lattice are close to one another, bound states of the corresponding types of solitary waves occur. Bound states of this sort are vector solitons, whose polarization varies along the pulse. If the transverse acoustic modes are degenerate in velocity, the situation is extremely similar to the propagation of pulses in optical fibers.

INTRODUCTION

This paper is devoted to a theoretical study of the dynamics of three-dimensional anharmonic crystal lattices in which only acoustic waves propagate in the harmonic interaction limit. The problem of realizing soliton states in anharmonic lattice systems has been studied in many papers.[1–8] For solitons whose dimensions exceed the lattice constant, this problem reduces to a nonlinear Schrödinger equation, so that this group of papers can be expanded to include studies of polaron motion in the adiabatic approximation.[9–14] In almost all of these papers, interest revolves around the model of a discrete nonlinear Schrödinger equation for one-dimensional chains. In the most interesting case of soliton states with large dimensions (compared to the lattice constant), the discreteness of the model should lead only to weak pinning of the solitons (cf. the discrete Frenkel–Kontorova model or the motion of Josephson vortices[15]), but has little effect on the shape of the solitons.

On the other hand, the problem becomes fundamentally different for three-dimensional lattices. In this case, a longitudinal and two transverse acoustic modes propagate in a monatomic crystal, whereas only longitudinal sound is present in one-dimensional chains. Thus when anharmonicity is taken into account in three-dimensional structures, we should expect the appearance of three types of soliton states, described by a system of three coupled nonlinear equations. This situation resembles the development of solitons in optical fibers. In optics, however, there is no longitudinal component of the electromagnetic oscillations, so the propagation of pulses in a fiber is described by a system of just two coupled nonlinear equations.[16] In the nonlinear optics of a birefringent fiber, in particular, this leads to the formation of vector soliton states whose polarization varies along the pulse.[17–22]

In the following, a system of nonlinear equations is derived for acoustic solitons in three-dimensional crystals and it is shown that under certain conditions the corresponding soliton states can have a vector character. As a matter of fact, the type of solitary waves in three-dimensional crystals is determined to a great extent by the symmetry of the crystal lattice and the propagation direction relative to the crystallographic axes.

NONLINEAR EQUATIONS OF MOTION IN THE CONTINUUM LIMIT

We limit ourselves below to the case of the continuum limit and, in order to obtain the equations of motion, we use the methods of the theory of elasticity.[23] To avoid cumbersome calculations, from the outset we assume that the displacement **u** of the medium depends only on a single coordinate x, so that the strain tensor

$$u_{ik}=1/2\,(v_i+v_k+\delta_{ik}v_l^2),$$

where $v_i=du_i/dx$. In this case, for a cubic crystal, if the x axis (the direction of motion) coincides with a crystallographic axis, then the energy of the medium, including a fourth-order anharmonicity, has the form

$$\Omega=\rho\int dx\left(\frac{1}{2}c_l^2v_x^2+\frac{1}{2}c_t^2v_t^2+\frac{1}{4}Av_x^4+\frac{1}{2}Bv_x^2v_t^2+\frac{1}{4}Cv_t^4\right), \tag{1}$$

where ρ is the density of the medium, $v_t^2=v_y^2+v_z^2$, c_l and c_t are, respectively, the longitudinal and transverse sound velocities, and A, B, and C are the fourth-order anharmonicity constants.

The third-order anharmonicity has been neglected in Eq. (1), since, as we shall see below, it does not contribute to the soliton states being studied here.

In our case, the equation of motion of the theory of elasticity, $\rho\ddot{u}_i=\partial\sigma_{ik}/\partial x_k$, where $\sigma_{ik}=\partial\Omega/\partial u_{ik}$ is the stress tensor, takes the form

$$\frac{\partial^2 u_x}{\partial t^2}=\frac{\partial}{\partial x}(c_l^2 v_x+Av_x^3+Bv_xv_t^2),$$

$$\frac{\partial^2 u_{y,z}}{\partial t^2}=\frac{\partial}{\partial x}(c_t^2 v_{y,z}+Bv_{y,z}v_x^2+Cv_{y,z}v_t^2).$$

On differentiating these equations with respect to *x*, we obtain the equation of motion for the quantity v_i:

$$\frac{\partial^2 v_x}{\partial t^2}=\frac{\partial^2}{\partial x^2}(c_l^2 v_x+Av_x^3+Bv_xv_t^2),$$

$$\frac{\partial^2 v_{y,z}}{\partial t^2}=\frac{\partial^2}{\partial x^2}(c_t^2 v_{y,z}+Bv_{y,z}v_x^2+Cv_{y,z}v_t^2). \qquad (2)$$

Equations of the form (2) are not sufficient for describing stable soliton states. This is well known in nonlinear optics. The anharmonic contribution to the equations of motion (for $A,B,C>0$) is responsible for pulse compression. In order to maintain stability it is necessary to include a mechanism for pulse spreading. In nonlinear optics one such mechanism is frequency dispersion of the dielectric constant.[16] In acoustics, an analogous role is played by the spatial dispersion of the sound velocity. In the language of the theory of elasticity, this means including a contribution to the energy (1) from the spatial derivatives $\partial v_i/\partial x$. With this generalization, the equations of motion (2) take the form

$$\frac{\partial^2 v_x}{\partial t^2}=\frac{\partial^2}{\partial x^2}\left(c_l^2 v_x+\alpha_l\frac{\partial^2 v_x}{\partial x^2}+Av_x^3+Bv_xv_t^2\right),$$

$$\frac{\partial^2 v_{y,z}}{\partial t^2}=\frac{\partial^2}{\partial x^2}\left(c_t^2 v_{y,z}+\alpha_t\frac{\partial^2 v_{y,z}}{\partial x^2}+Bv_{y,z}v_x^2+Cv_{y,z}v_t^2\right). \qquad (3)$$

Here α_l and α_t are constants describing the velocity dispersion of the longitudinal and transverse sound, respectively. On going to the model of a one-dimensional chain ($v_y=v_z=0$), Eq. (3) takes the form of a nonlinear modified Boussinesq equation for v_x (Ref. 1). This equation has an exact solution in the form of a soliton propagating at a velocity *V* exceeding the longitudinal sound velocity,

$$v_x(x,t)=\left[\frac{2(V^2-c_l^2)}{A}\right]^{1/2}\cosh^{-1}\left[\frac{x-Vt+x_0}{L}\right],$$

where the soliton size is $L=\sqrt{\alpha_l/(V^2-c_l^2)}$.

A search for solutions of the modified Boussinesq equation in the form of a pulse with a high carrier frequency leads to a nonlinear Schrödinger equation for the envelope.[1] In essence, the following calculations are a generalization of the construction of these solutions for solitary waves in the three-dimensional case.

The form of the equations of motion (3) suggests a close analogy between the nonlinear acoustics problem being studied here and the problem of the propagation of light pulses in optical fibers. In fact, if we interchange the coordinate and time, i.e., $x\leftrightarrow t$, and take v_i to be the projection of the electric field **E** of the wave, then the quantities in parentheses in Eq. (3) can be identified with the electric displacement **D**. Here the sound velocities play the role of the linear permittivity, and the spatial dispersion of the sound velocity plays the role of the frequency dispersion of the permittivity, etc. As a result, this equation takes the form of the wave equation $\partial^2\mathbf{E}/\partial x^2-c^{-2}\partial^2\mathbf{D}/\partial t^2=0$. There is, however, an important difference between the nonlinear acoustics and optics problems. It lies in the fact that the electromagnetic wave has no longitudinal component ($\mathbf{E}_x=0$), while all three components of the ''vector'' **v** exist in acoustics.

A simplification of the system of Eqs. (3) can be achieved by assuming that the variables v_i are modulated by some carrier frequency ω, so that $v_i=a\exp(-i\omega t)+a^*\exp(i\omega t)$. We shall leave aside the problems associated with the generation of overtones at frequencies $n\omega$ and only consider oscillations at the fundamental carrier frequency ω. In this approximation, on going to the complex representation $v_i=a_i\exp(-i\omega t)$, we can make the substitution $v_i^3\rightarrow v_i|v_i|^2$, $v_iv_k^2\rightarrow v_i|v_k|^2+(1/2)v_i^*v_k^2$ in Eq. (3). As a result, Eq. (3) takes the form

$$\frac{\partial^2 v_x}{\partial t^2}=\frac{\partial^2}{\partial x^2}\left(c_l^2 v_x+\alpha_l\frac{\partial^2 v_x}{\partial x^2}+Av_x|v_x|^2 + B\left[v_x|v_t|^2+\frac{1}{2}v_x^*v_t^2\right]\right),$$

$$\frac{\partial^2 v_{y,z}}{\partial t^2}=\frac{\partial^2}{\partial x^2}\left(c_t^2 v_{y,z}+\alpha_t\frac{\partial^2 v_{y,z}}{\partial x^2}+B\left[v_{y,z}|v_x|^2+\frac{1}{2}v_{y,z}^*v_x^2\right] + C\left[v_{y,z}|v_t|^2+\frac{1}{2}v_{y,z}^*v_t^2\right]\right). \qquad (4)$$

In connection with the transition from Eq. (3) to the complex form (4) and leaving out the overtone contributions, it is clear that the third-order anharmonicity which was omitted before makes no contribution to the effects being studied here, since it contains no terms that are modulated at the fundamental carrier frequency.

SYSTEM OF NONLINEAR SCHRöDINGER EQUATIONS

Before proceeding to a transformation of Eqs. (4), let us generalize them to the case of a weak anisotropy in the *y*, *z* plane. Suppose that the velocities c_i for the polarizations of transverse sound in the *y* and *z* directions differ slightly. Such a difference might arise, for example, in a cubic crystal when a pressure is applied along the *y* axis. By analogy with light, this corresponds to a birefringent medium. Here, as in problems involving birefringence, we neglect the anisotropy of the anharmonic terms. This generalization reduces to the substitution $c_t\rightarrow c_y$ in Eq. (4) for v_y and $c_t\rightarrow c_z$ in the equation for v_z. We shall assume that the anisotropy is weak, so that $|c_y-c_z|\ll c_t$, where now $c_t=(c_y+c_z)/2$ is the average velocity. After this generalization, we shall seek a solution of Eq. (4) in the form

$$v_x(x,t)=a_x(x,t)\exp[i(q_lx-\omega t)],$$

$$v_{y,z}(x,t)=a_{y,z}(x,t)\exp[i(q_tx-\omega t)]. \qquad (5)$$

The wave vectors q_l (for the longitudinal component) and q_t (for the transverse components) are determined by the value of the carrier frequency ω, according to the dispersion relations

$$\omega^2=c_l^2q_l^2-\alpha_l q_l^4; \quad \omega^2=c_t^2q_t^2-\alpha_t q_t^4. \tag{6}$$

We assume that the amplitudes $a_i(x,t)$ vary slowly in space (over distances of order q^{-1}) and time (over intervals of order ω^{-1}). Substituting Eq. (5) in Eq. (4) and leaving out the terms proportional to $\partial^4 a_i/\partial x^4$ and $\partial^3 a_i/\partial x^3$, as well as all the spatial derivatives in the anharmonic terms, we obtain

$$\begin{aligned}\frac{\partial^2 a_x}{\partial t^2}-2i\omega\frac{\partial a_x}{\partial t}={}&c_l^2\left(\frac{\partial^2 a_x}{\partial x^2}+2iq_l\frac{\partial a_x}{\partial x}\right)-\alpha_l\left(6q_l^2\frac{\partial^2 a_x}{\partial x^2}\right.\\&\left.+4iq_l^3\frac{\partial a_x}{\partial x}\right)-Aq_l^2a_x|a_x|^2\\&-Bq_l^2a_x(|a_y|^2+|a_z|^2)-\frac{B}{2}(2q_t\\&-q_l)^2a_x^*(a_y^2+a_z^2)\exp[2i(q_t-q_l)x],\\\frac{\partial^2 a_y}{\partial t^2}-2i\omega\frac{\partial a_y}{\partial t}={}&c_y^2\left(\frac{\partial^2 a_y}{\partial x^2}+2iq_t\frac{\partial a_y}{\partial x}\right)+q_t^2(c_z^2-c_y^2)a_y\\&-\alpha_t\left(6q_t^2\frac{\partial^2 a_y}{\partial x^2}+4iq_t^3\frac{\partial a_y}{\partial x}\right)\\&-Cq_t^2\left[a_y(|a_y|^2+|a_z|^2)+\frac{1}{2}a_y^*(a_y^2\right.\\&\left.+a_z^2)\right]-Bq_t^2a_y|a_x|^2-\frac{B}{2}(2q_l\\&-q_t)^2a_y^*a_x^2\exp[2i(q_l-q_t)x].\end{aligned}\tag{7}$$

The equation for a_z can be obtained from that for a_y by making the substitution $y\leftrightarrow z$. The system of three nonlinear equations (Eq. (7) plus that for a_z) can be simplified. This simplification is related to the fact that in crystals with cubic symmetry the velocities of the longitudinal and transverse sound waves are substantially different. (Usually $c_l^2>2c_t^2$.) Therefore pulses polarized along the x axis propagate much faster than pulses polarized in the yz plane. Thus they cannot create a bound state, since the solitary waves realized in this system are polarized either along the x axis or perpendicular to it. One can only speak of their interaction in collisions. The situation is analogous to solitons in highly birefringent optical fibers, where solitons polarized along the fast and slow axes do not form bound states but undergo radiative decay in collisions.[24] (Collisions between acoustic solitons in one-dimensional chains have been studied numerically.[1]) As we are not interested in this sort of collisions, which are described in Eq. (7) by the anharmonic constant B, we can set $a_y=a_z=0$ in the equation for a_x (or, equivalently, $B=0$). As a result, we obtain a closed equation for the amplitude of a solitary longitudinally polarized wave. We now transform in this equation to a moving coordinate system with $a_x(x,t)\rightarrow a_x(\zeta,t)$, where

$$\zeta=x-\frac{d\omega}{dq_l}t.$$

The propagation velocity equals the group velocity for longitudinal sound determined by the dispersion relation (6). As a result, after some simple transformations, Eq. (7) yields

$$i\frac{\partial a_x}{\partial t}=\beta_l\frac{\partial^2 a_x}{\partial\zeta^2}+\gamma_l a_x|a_x|^2, \tag{8}$$

where

$$\beta_t=\frac{3\alpha_l\omega}{2c_l^2}, \quad \gamma_l=\frac{\omega A}{2c_l^2}. \tag{9}$$

In deriving Eq. (10), we have, as usual, left out the small contributions proportional to $\partial^2 a_x/\partial t^2$ and $\partial^2 a_x/\partial t\partial\zeta$.

The scalar nonlinear Schrödinger equation (8) describes longitudinally polarized solitary waves and is identical to the analogous equation for one-dimensional chains.[1] It can be integrated by the techniques used for inverse scattering problems and has been studied in many papers. Thus, in the following we shall not dwell on it. We now proceed to transversely polarized acoustic solitons. To do this, we set $a_x=0$ in the second of Eqs. (7) and again transform to a coordinate system moving at the group velocity of transverse sound, i.e., $a_{y,z}(x,t)\rightarrow a_{y,z}(\xi,t)$, where $\xi=x-(d\omega/dq_t)t$. Through transformations similar to those used in deriving Eq. (8), we obtain a system of two equations

$$\begin{aligned}i\frac{\partial a_y}{\partial t}={}&\kappa a_y+i(c_z-c_y)\frac{\partial a_y}{\partial\xi}+\beta_t\frac{\partial^2 a_y}{\partial\xi^2}\\&+\gamma_t\left\{a_y\left(|a_y|^2+\frac{2}{3}|a_z|^2\right)+\frac{1}{3}a_y^*a_z^2\right\},\\i\frac{\partial a_z}{\partial t}={}&-\kappa a_y-i(c_z-c_y)\frac{\partial a_z}{\partial\xi}+\beta_t\frac{\partial^2 a_z}{\partial\xi^2}\\&+\gamma_t\left\{a_z\left(|a_z|^2+\frac{2}{3}|a_y|^2\right)+\frac{1}{3}a_z^*a_y^2\right\}.\end{aligned}\tag{10}$$

Here

$$\beta_t=\frac{3\alpha_t\omega}{2c_t^2}, \quad \gamma_t=\frac{3\omega C}{4c_t^2},$$

$$\kappa=\frac{\omega}{2}\frac{c_z^2-c_y^2}{c_z^2+c_y^2}\cong\frac{\omega}{2}\frac{c_z-c_y}{c_t}. \tag{11}$$

For concreteness, in the following we assume that $c_z>c_y$, i.e., $\kappa>0$.

After the substitution $\xi\leftrightarrow t$, the system of Eqs. (10) is outwardly the same as the equations for solitary waves in birefringent optical fibers,[16] so the analysis employed below is analogous to that proposed elsewhere[21] for studying vector solitons in nonlinear optics.

VECTOR ACOUSTIC SOLITONS

Equations (10) differ from the standard nonlinear Schrödinger equations in having terms proportional to $\partial a_{y,z}/\partial\xi$. They occur with different signs in the equations for a_y and a_z because the corresponding solitons propagate at different velocities owing to the difference in the group velocities of waves polarized along the x and y axes. Because of the small difference in these velocities, however, there is some hope that bound states can be realized. Let us make a transformation which makes it possible to remove these first derivatives with respect to the coordinates from the equations,

$$a_{y,z}(\xi,t)=b_{y,z}(\xi,t)\exp\left\{ i\,\frac{(c_z-c_y)}{4\beta_t}\,t\mp i\Lambda\xi\right\}, \tag{12}$$

where the spatial frequency

$$\Lambda=\frac{c_z-c_y}{2\beta_t}. \tag{13}$$

As a result, Eqs. (10) take the form

$$i\frac{\partial b_y}{\partial t}=\kappa b_y+\beta_t\frac{\partial^2 b_y}{\partial\xi^2}+\gamma_t\left\{ b_y\left(|b_y|^2+\frac{2}{3}|b_z|^2\right)\right.$$
$$\left.+\frac{1}{3}b_y^* b_z^2\exp(4i\Lambda\xi)\right\},$$
$$i\frac{\partial b_z}{\partial t}=-\kappa b_z+\beta_t\frac{\partial^2 b_z}{\partial\xi^2}+\gamma_t\left\{ b_z\left(|b_z|^2+\frac{2}{3}|b_y|^2\right)\right.$$
$$\left.+\frac{1}{3}b_z^* b_y^2\exp(-4i\Lambda\xi)\right\}. \tag{14}$$

In Eq. (14), as opposed to Eq. (10), the coefficients depend explicitly on the coordinate and are invariant under the substitution $\xi\rightarrow\xi+\pi/2\Lambda$. Thus we might expect that the solution is a set of periodically positioned pulses, separated by distances that are multiples of the length $\pi/2\Lambda$. If the length λ_i of a pulse is much shorter than this distance, then neighboring pulses overlap weakly. Thus, in order to find an approximate equation for a single pulse we can replace the factors $\exp(\pm 4i\Lambda\xi)$ in Eq. (14) by $\exp(\pm 4i\Lambda\xi_0)$, where ξ_0 is the location of the center of gravity of the pulse, provided, of course, that $\lambda_t\Lambda\ll 1$. After this approximation, we seek a solution of the form

$$b_{y,z}(\xi,t)=f_{y,z}(\xi)\exp(-iQt\mp i\Lambda\xi_0), \tag{15}$$

where $f_{y,z}(\xi)$ are real amplitudes.

As a result, Eq. (14) takes the form

$$\beta_t\frac{d^2 f_y}{d\xi^2}=(Q-\kappa)f_y-\gamma_t f_y(f_y^2+f_z^2),$$
$$\beta_t\frac{d^2 f_z}{d\xi^2}=(Q+\kappa)f_z-\gamma_t f_z(f_y^2+f_z^2). \tag{16}$$

Equations (15) are formally the same as Newton's equations of motion for a particle of mass β_t in a two-dimensional space f_y, f_z with potential energy

$$U=-\tfrac{1}{2}(Q-\kappa)f_y^2-\tfrac{1}{2}(Q+\kappa)f_z^2+\tfrac{1}{4}\gamma_t(f_y^2+f_z^2)^2. \tag{17}$$

The solution of Eqs. (16) can be sought by separation of variables in an elliptical coordinate system and using the formalism of the Hamilton–Jacobi method. Here it is necessary to find all possible trajectories of the particle motion corresponding to solitary waves. Such a procedure has been carried out in Ref. 21, so here we can make direct use of the results from that paper.

Ordinary ''bright'' solitons correspond to trajectories which begin and end at the coordinate origin $f_y=f_x=0$, which corresponds to having the pulse amplitude go to zero at $\xi\rightarrow\pm\infty$. The amplitude of a bright soliton is given by

$$f_y=\pm 2\sqrt{\frac{\kappa}{\gamma_l}}$$
$$\times\frac{\eta_-\sinh[\eta_+(\xi-\xi_0)-\delta]}{\eta_+\cosh[\eta_+(\xi-\xi_0)-\delta]\cosh[\eta_-(\xi-\xi_0)-\delta]-\eta_-\sinh[\eta_+(\xi-\xi_0)-\delta]\sinh[\eta_-(\xi-\xi_0)-\delta]},$$

$$f_z=2\sqrt{\frac{\kappa}{\gamma_l}}$$
$$\times\frac{\eta_+\cosh[\eta_-(\xi-\xi_0)-\delta]}{\eta_+\cosh[\eta_+(\xi-\xi_0)-\delta]\cosh[\eta_-(\xi-\xi_0)-\delta]-\eta_-\sinh[\eta_+(\xi-\xi_0)-\delta]\sinh[\eta_-(\xi-\xi_0)-\delta]}, \tag{18}$$

where $\eta\pm=(Q\pm\kappa)/\beta_t$, δ is an arbitrary parameter which determines the set of different vector solitons, and ξ_0 is an arbitrarily chosen coordinate origin.

We shall refer to these as vector solitons because their polarization vector varies along the pulse. According to Eq. (18), the polarization angle Θ of a soliton varies with the coordinate as

$$\tan\Theta=\frac{f_y}{f_z}=\pm\frac{\eta_-}{\eta_+}\,\frac{\sinh(\eta_+\xi-\delta)}{\cosh(\eta_-\xi-\delta)}. \tag{19}$$

Here the coordinate origin is $\xi_0=0$. Thus the polarization vector varies from $\mp\pi/2$ for $\xi\rightarrow-\infty$ to $\pm\pi/2$ for $\xi\rightarrow\infty$, passing through zero at the point $\xi=\delta/\eta_+$.

According to Eq. (18), a vector soliton is a superposition of two pulses polarized along the y and z axes and whose centers of gravity are displaced relative to one another by a distance proportional to the parameter δ. In the degenerate limit where the transverse sound velocity $c_y=c_z$ ($\kappa=0$, $\eta_+=\eta_-$), this parameter corresponds to a scalar soliton polarized in the yz plane. In the isotropic limit, of course, the soliton does not have a vector character, since its polarization does not change along the pulse and the pulse shape is independent of the polarization angle. In the special case of $\kappa=0$, we have $f_y=f\sin\Theta$ and $f_z=f\cos\Theta$, with

$$f=\sqrt{\frac{2Q}{\gamma_t}}\cosh^{-1}\left((\xi-\xi_0)\sqrt{\frac{Q}{\beta_t}}\right). \tag{20}$$

One interesting feature of the vector soliton states obtained here is that the elastic crystalline energy stored in them is invariant with respect to the parameter δ. Using Eq. (18), we find that

$$W=\int_{-\infty}^{\infty} d\xi (f_y^2+f_z^2)=\frac{\sqrt{4\beta_t\kappa}}{\gamma_t}\left\{\sqrt{\frac{Q}{\kappa}+1}+\sqrt{\frac{Q}{\kappa}-1}\right\}. \tag{21}$$

Thus, as in the case of ordinary solitons described by a scalar Schrödinger equation (8), the parameter Q determines the power in the pulse. Vector solitons, however, can exist only in regions with sufficiently high Q such that $Q>\kappa$. Thus there is a threshold power $W_{\min}=\sqrt{8\kappa\beta_t}/\gamma_t$ at $Q=\kappa$ for vector solitons, below which these states are not realized.

For $Q<\kappa$, soliton states develop which are polarized strictly along the z axis. In this region, $f_y=0$ and

$$f_z=\sqrt{\frac{2(Q+\kappa)}{\gamma_t}}\cosh^{-1}\left((\xi-\xi_0)\sqrt{\frac{Q+\kappa}{\beta_t}}\right). \tag{22}$$

A potential energy in the form (17) permits the existence of yet another trajectory corresponding to a solitary wave. This trajectory joins the points $f_y=-\sqrt{(Q-\kappa)/\gamma_t}$, $f_z=0$ and $f_y=\sqrt{(Q-\kappa)/\gamma_t}$, $f_z=0$, and passes through the absolute equilibrium point $f_y=0$, $f_z=\pm\sqrt{(Q+\kappa)/\gamma_t}$. The corresponding solution has the character of a mixed soliton: its projection along the f_z axis is a ''bright'' soliton (with amplitude going to zero at $\xi\to\pm\infty$), while the second component f_y is a ''dark'' soliton with nonzero amplitude at $\xi\to\pm\infty$:

$$f_y=\sqrt{\frac{Q-\kappa}{\gamma_t}}\tanh\left((\xi-\xi_0)\sqrt{\frac{\kappa}{\beta_t}}\right),$$

$$f_z=\sqrt{\frac{Q+3\kappa}{\gamma_t}}\cosh^{-1}\left((\xi-\xi_0)\sqrt{\frac{\kappa}{\beta_t}}\right). \tag{23}$$

Mixed solitons of this type can exist only in the region $Q>\kappa$. The ''dark'' component of the soliton essentially describes a change by π in the phase of the transverse sound polarized along the y axis. This phase change is accompanied by an acoustic pulse of the orthogonal polarization along the z axis. Note that similar mixed soliton states have been studied in nonlinear optics (Ref. 18; see also Ref. 21).

DISCUSSION OF RESULTS

It has been shown above that three-dimensional crystals support three types of solitary waves, which are described by three coupled nonlinear Schrödinger equations (7) in accordance with the three branches of acoustic phonons present in the harmonic approximation. The specific form of the solitary waves is determined to a substantial extent by the type of crystal lattice and the propagation direction. Thus, in biaxial crystals, where the group velocities of all three phonon modes differ greatly, vector solitons are unlikely to develop. In these lattices we may expect the appearance of ordinary soliton pulses of three types, each of which is polarized along one of the principal axes and moves at its own velocity, equal to the group velocity of the corresponding acoustic mode. In uniaxial crystals for light propagating along the c axis or in cubic crystals, the velocities of the transverse acoustic modes are degenerate and differ greatly from the velocity of the longitudinal oscillations. Thus these lattices should support ordinary solitons polarized along the propagation direction and vector solitons polarized in the perpendicular plane. In fact, these solitons will also have a vector character if the equality of the group velocities for transverse sound is slightly disturbed by some external interaction. Because there is no interaction between the longitudinal and transverse solitons, the situation for transverse pulses is extremely similar to that which exists in studies of optical solitons propagating in fibers. An exception may be the special case of a uniaxial crystal in which the longitudinal and transverse sound velocities are close to one another. In this case, three-dimensional vector solitons may develop in which the polarization vector rotates in two directions along a pulse.

The formation of a bound state of the vector soliton type is based on the idea that a small difference in the group velocities of sound for different polarizations is compensated by the difference in the difference Λ in the wave vectors of the corresponding carrier frequencies (see Eq. (12)). In what follows we make the approximate substitution $\exp(\pm 4i\Lambda\xi)\to\exp(\pm 4i\Lambda\xi_0)$ in the multipliers in Eq. (14), making use of the smallness of the pulse length λ_i compared to Λ^{-1}. An analysis shows that including corrections in the small parameter $\lambda_i\Lambda$ leads to the emission of acoustic waves by a moving vector soliton (i.e., a pedestal). This emission causes a gradual decay of the pulse state. Stability of the system can be recovered by ensuring propagation of a periodic sequence of pulses separated by a distance $\pi/2\Lambda$ and modulated by a synchronized carrier frequency. Then the energy of the accompanying acoustic background (pedestal) is exchanged among neighboring pulses and destruction of the soliton states ceases. In optics, the passive synchronization of a periodic sequence of pulses in fiber optic lasers may be based on this effect.[24,25]

Numerical studies of the propagation of acoustic pulses in anharmonic lattices are now actively under way.[1,4–8] In particular,[1,4,6] acoustic pulses have been observed to propagate in one-dimensional chains without significant change in their shapes; this was attributed to the development of soliton states of various types. However, a reliable identification of soliton states requires studies of the long-time evolution of the pulse states. There have been no studies, at all, of the spontaneous formation of solitons from a noise background. Thus, no data are currently available on the characteristic formation length for acoustic solitons. This situation is made more complicated by the fact that anharmonic lattices can in principle support solitary waves of different types, with different propagation velocities. In the present paper, primary attention has been devoted to solitons which propagate at the group velocity of sound and obey the nonlinear Schrödinger equation. The more general Boussinesq equation (see Eq. (3)), however, admits solutions in the form of solitons propagating at a supersonic velocity V and having the form of a pulse without modulation by a carrier frequency. As V approaches the sound velocity, the Boussinesq equation can be reduced to the form of the Korteweg–de Vries equation, which has exact solutions in the form of solitons of an oscillatory type (breathers). In three-dimensional lattices the situation is still more complicated, and systems of this sort have not been studied yet, even numerically.

This work was supported by the Russian Fund for Fundamental Research, Grant No. 96-02 16848-a.

[1] N. Flytzanis, St. Pnevmatikov, and M. Remoissenet, J. Phys. C **18**, 4603 (1985).

[2] Yu. S. Kivshar, Phys. Rev. B **48**, 4132 (1993).

[3] S. Takeno, Prog. Theor. Phys. **71**, 395 (1984).

[4] G. S. Zavt, M. Wagner, and A. Luetze, Phys. Rev. B **47**, 4108 (1993).

[5] M. Wagner, G. Zavt, J. Vazquez-Marquez *et al.*, Philos. Mag. B **65**, 273 (1992).

[6] B. Lindquist, M. Johansson, and R. Riklund, Phys. Rev. B **50**, 9860 (1994).

[7] G. Kalosakas, G. P. Tsironis, and E. N. Economou, J. Phys. **6**, 7847 (1994).

[8] J. Vasquez-Marquez, M. Wagner, M. Montagna *et al.*, Phys. B **172**, 355 (1991).

[9] S. Takeno, J. Phys. Soc. Jpn. **59**, 3127 (1990).

[10] X. Wang, D. W. Brown, and K. Lindenberg, Phys. Rev. B **39**, 5366 (1989).

[11] G. Kopidakis, C. M. Soukoulis, and E. N. Economou, Phys. Rev. B **49**, 7036 (1994).

[12] O. M. Braun and Yu. S. Kivshar, Phys. Rev. B **50**, 13388 (1994).

[13] A. V. Zolotaryuk, K. M. Spatschek, and O. Kluth, Phys. Rev. B **47**, 7827 (1993).

[14] D. Chen, M. I. Molina, and G. P. Tsironis, J. Phys. **5**, 8689 (1993).

[15] V. V. Bryksin and S. N. Dorogovtsev, Zh. Éksp. Teor. Fiz. **102**, 1025 (1992) [Sov. Phys. JETP **75**, 558 (1992)].

[16] G. P. Agrawal, *Nonlinear Fiber Optics*, Academic Press, Boston (1989).

[17] D. N. Christodoulides and R. I. Joseph, Opt. Lett. **13**, 53 (1988).

[18] D. N. Christodoulides, Phys. Lett. **132**, 451 (1988).

[19] M. N. Islam, C. R. Menyuk, C. J. Chen, and C. E. Soccolich, Opt. Lett. **16**, 214 (1991).

[20] V. V. Bryksin, M. P. Petrov, and R. V. Kiyan, Pis'ma Zh. Tekh. Fiz. **20**(10), 6 (1994) [Tech. Phys. Lett. **20**, 386 (1994)].

[21] V. V. Bryksin, M. P. Petrov, and R. V. Kiyan, Zh. Éksp. Teor. Fiz. **107**, 732 (1995) [JETP **80**, 415 (1995)].

[22] C. De Angelis, M. Santagiustina, and S. Wabnitz, Opt. Commun. **122**, 23 (1995).

[23] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 2nd ed., [Pergamon Press, Oxford (1970); Nauka, Moscow (1965)].

[24] B. A. Malomed, Phys. Rev. A **43**, 410 (1991).

[25] V. V. Bryksin and M. P. Petrov, Pis'ma Zh. Tekh. Fiz. **22**(4), 46 (1996) [Tech. Phys. Lett. **22**, 153 (1996)].

Translated by D. H. McNeill

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Influence of the thickness of the liquid layer on the ratio of the dimensions of a convection cell

E. D. Éĭdel'man

St. Petersburg Chemical-Pharmaceutical Institute, St. Petersburg, Russia

(Submitted May 12, 1996)
Zh. Tekh. Fiz. **68**, 7–11 (November 1998)

A theory is developed for a new type of transition — a change in the ratio of the longitudinal and transverse dimensions of a convection cell as the thickness of a liquid layer is varied. A sudden change in the ratio of the cell dimensions takes place because of a change in the predominant mechanism for excitation of convection. The governing influence of buoyancy forces gives way to one of thermocapillary forces, and they in turn give way to the influence of thermoelectric forces for yet thinner layers. As the layer thickness is reduced gradually at a fixed external heating, the ratio of the dimensions will take on the values 0.7, 0.65, and 1, respectively. © *1998 American Institute of Physics.* [S1063-7842(98)00211-6]

INTRODUCTION

Three mechanisms for the excitation of thermal convection with heating from below are presently known. First, there is the Rayleigh mechanism. This mechanism is comes into being through the buoyancy force, i.e., the difference between the Archimedean force and the force of gravity.[1] Second, there is the mechanism discovered by Pearson.[2] This mechanism due to the difference in the surface tension forces arising from the temperature dependence of the surface tension forces, i.e., the thermocapillary effect.[3] Finally, the third mechanism[4] is associated with the excitation of motion by an electric force. This mechanism predominates in thin layers of liquid semiconductors (semimetals), electrolytes, and some other media. Heating of these media creates a thermoelectric field. This field acts on the charge fluctuations which arise for the same reason, and the result is a Coulomb force which sets the liquid into motion.[5]

All three of these excitation mechanisms have been well studied, both separately (i.e., Rayleigh,[1] thermocapillary,[6], and thermoelectric[7]) and in pairs (i.e., the combined effects of buoyancy and thermocapillary forces,[8] buoyancy and thermoelectric forces,[9] and the combined action of forces owing to thermocapillarity and thermoelectricity[10]), and even when all three mechanisms act together.[9] It turns out that each of these excitation mechanisms has a predominant effect on the liquid within some range of parameters. It is important that in a given liquid and for sufficient external heating, the sole parameter determining which excitation mechanism will make the dominant contribution to the motion of the liquid is the thickness of the liquid layer.

In this paper we concentrate on the features of the transition from one excitation mechanism to another, whereas in the past the problems have been solved under conditions such that one of the mechanisms predominates, while the other (others) only cause small changes in its effect.

1. CONDITIONS FOR MAINTAINING THE RATIO OF THE DIMENSIONS OF A CONVECTION CELL

Each of the excitation mechanisms mentioned above can actually set a liquid medium into motion if only the driving force exceeds the dissipative force $\rho\nu\varkappa$, where ρ is the liquid density, while ν and $\varkappa$ are, respectively, the dissipative coefficients of kinematic viscosity and thermal diffusivity.

Under a heating $A=(T_h-T_c)/h$, which is the difference between the temperatures T_h of the hot (lower) and T_c of the cold (upper) surfaces of the layer divided by the layer thickness h, the following forces act on the liquid:

1. Buoyancy force $\rho\beta gAh^4$ (Ref. 1). This force is created by the volume expansion of the liquid owing to heating (β is the coefficient of thermal expansion and g is the acceleration of gravity).
2. Thermocapillary force σAh^2 (Ref. 3). This force is created by the thermocapillarity effect (σ is the coefficient of thermocapillarity).
3. Thermoelectric force $\varepsilon\gamma^2A^2h^2$ (Ref. 5). This force is created by the thermoelectric effect (γ is the thermopower (Siebeck coefficient) in a liquid with a dielectric permittivity ε high enough to keep the space charge which develops in the liquid from dissipating.

The action of each mechanism is characterized by a dimensionless number. These numbers are, respectively, the Rayleigh number $\mathcal{R}$, the Marangoni number $\mathcal{M}$, and the ''thermoelectric number'' $\mathcal{E}=I^2$. We have

$$\mathcal{R}=\frac{\beta gAh^4}{\nu\varkappa},\quad \mathcal{M}=\frac{\sigma Ah^2}{\rho\nu\varkappa},\quad \mathcal{E}=I^2=\frac{\varepsilon\gamma^2A^2h^2}{\rho\nu\varkappa}. \tag{1}$$

These numbers represent the factor by which the force exciting the motion exceeds the dissipative force. When at least one of the dimensionless numbers reaches a critical value ($\mathcal{R}^*>660$, $\mathcal{M}^*>80$, $I^*>6.3$), motion develops in the liquid, even if the other exciting forces do not act. The specific values of $\mathcal{R}^*$, $\mathcal{M}^*$, or I^* required for an instability to set in depend on the boundary conditions, e.g., on the heat transfer conditions. The minimum values of 660, 80, and 6.3, respectively, are sufficient for excitation when the boundaries are held at strictly constant temperatures.

Motion develops at the minimum possible values of these dimensionless numbers for the given boundary conditions. The minimum condition determines the ratio of the dimensions of the convection cell that develops at the moment of excitation. Denoting the dimension of the cell along the layer by l (the perpendicular dimension is always equal to the layer thickness h), we find that for excitation by the buoyancy force, thermocapillarity, or thermoelectricity, the ratio l^2/h^2 will be equal to 0.5, 0.45, or 1, respectively. It turns out that the ratio l/h[1,3,9] for a given excitation mechanism is independent of the boundary conditions. Thus the ratio of the dimensions of a convection cell along and perpendicular to the layer can serve as an indicator of the excitation mechanism.

It is known[1] that the ratio of the dimensions of an emerging cell is retained in developed convection. At least, this is true as long as the amplitudes of the velocity and other convective quantities are proportional to the square root of the supercriticality, $(T_h - T_h^*)^{1/2}$ (Ref. 11; T_h is the actual temperature of the heated lower surface and T_h^* is the temperature of that surface which is sufficient to excite the instability, with $T_h > T_h^*$).

For the different excitation mechanisms the amplitudes are proportional to $(\mathcal{R}-\mathcal{R}^*)^{1/2}$, $(\mathcal{M}-\mathcal{M}^*)^{1/2}$, and $(I-I^*)^{1/2}$, respectively.

2. QUALITATIVE STUDY OF THE CONDITIONS FOR A CHANGE IN THE RATIO OF THE DIMENSIONS OF A CONVECTION CELL

The range of variation in the layer thickness over which one or another excitation mechanism predominates can be obtained by analyzing the dependence of the dimensionless numbers $\mathcal{R}$, $\mathcal{M}$, and $\mathcal{E}$ on h. It is evident that $\mathcal{R} \propto h^3$ and $\mathcal{M} \propto h$, while $\mathcal{E}$ (and I) is independent of h. It follows immediately that the Rayleigh excitation mechanism should be predominant in the thickest films, the thermocapillarity effect should exert the main influence in films of intermediate thickness, and in the thinnest films the excitation occurs under the influence of the thermoelectric field. More-exact estimates can be made by comparing the numbers $\mathcal{R}$, $\mathcal{M}$, and $\mathcal{I}$.

A comparison of excitation by the buoyancy and thermocapillarity forces shows that the thermocapillarity mechanism predominates in thin layers with thicknesses

$$h < h_{RM} \simeq \left(\frac{\sigma}{\rho g \beta} \right)^{1/2}. \tag{2}$$

On comparing the conditions for excitation by the buoyancy and thermoelectric forces, we see that the thermoelectric mechanism predominates in thin layers with thicknesses

$$h < h_{RI} \simeq \left(\frac{\varkappa \nu \varepsilon \gamma^2}{\rho \beta^2 g^2} \right)^{1/6}. \tag{3}$$

Finally, comparing the conditions for excitation by the thermocapillarity and thermoelectric mechanisms, we find that the thermoelectric mechanism predominates in thin layers with thicknesses

$$h < h_{MI} \simeq h_c \simeq \left(\frac{\rho \varkappa \nu \varepsilon \gamma^2}{\sigma^2} \right)^{1/2}. \tag{4}$$

All three effects occur in liquid semiconductors (semimetals). Typical values of the parameters of liquids[12,13] can be taken for them: $\rho = 1-10$ g/cm^3, $\beta = 9-6 \times 10^{-4}$ K^{-1}, $\nu \simeq \varkappa \simeq 5 \times 10^{-2} - 1$ mm^2/s, and $\sigma = 1-3 \times 10^{-5}$ N/mK. For estimates we can take the temperature difference $T_h - T_c \simeq Ah \simeq 10^3 - 10^4$ K. In order to make an estimate, the thermopower γ must be known. It was not possible to find values of γ at the melting temperature and above in the literature. It is known,[14] however, that the electrical conductivity and thermoelectric coefficients do not change significantly at the melting point. In solid semiconductors above the Debye temperature T_D, the thermopower is governed by an entrainment effect and is given by

$$\gamma \simeq f \, \frac{k_B}{e} \, \frac{T_D}{T}, \tag{5}$$

where k_B is Boltzmann's constant and e is the carrier charge.

The constant f is such that γ is ≈ 100 μV/K.[15]

Substituting the numerical values in Eqs. (2)–(4), we find that

$$10 < h_c < 100 \ \mu\text{m}, \quad 1 < h_{RM} < 10 \text{ mm}. \tag{6}$$

Thus, in layers up to 10 μm thick (and possibly up to 100 μm), excitation by the thermoelectric force predominates and a cell with $l \simeq h$ develops. In layers with thickness from 10 μm (possibly 100 μm) to 1 mm (perhaps 10 mm), the main mechanism for excitation is through the thermocapillary force. Then a cell with $l \simeq 0.65h$ develops. Finally, in still thicker layers, a cellular motion with a size ratio $l \sim 0.7h$ develops under the influence of buoyancy forces.

3. STATEMENT OF THE PROBLEM OF CALCULATING THE RATIO OF THE DIMENSIONS OF CONVECTION CELLS. NUMERICAL SOLUTIONS

The simplest approach for finding sufficient conditions for the instability owing to the growth of the small perturbations is to search for nontrivial solutions of systems of homogeneous differential equations, linearized with respect to the small perturbations, subject to homogeneous boundary conditions. The existence of a range of the parameters for which these solutions exist indicates the possibility of a spontaneous transition of the medium into a new state; if these values of the parameters are physically possible, then an instability develops. This approach has been extremely successful for analyzing the buoyancy force.[1]

In general, one poses the problem of finding conditions for the existence of nontrivial solutions of the linearized system formed from the equations of motion, continuity, and heat transport of an incompressible liquid, together with the continuity equation for the current and the equations of electrostatics,

$$-\nu\Delta\mathbf{v}+\frac{1}{\rho}\nabla p_1+\beta T_1\mathbf{g}+\frac{e\gamma}{\rho}n_1\nabla T_0=0, \tag{7}$$

$$\text{div }\mathbf{v}=0, \tag{8}$$

$$-\varkappa\Delta T_1+(\mathbf{v}\cdot\nabla)T_0=0, \tag{9}$$

$$\text{div }(\mathbf{E}_1-\gamma\nabla T_1)=0, \tag{10}$$

$$\text{div }\mathbf{E}_1=\frac{e}{\varepsilon}n_1, \tag{11}$$

where Δ denotes the Laplacian operator, $\mathbf{v}$ is the velocity of the liquid, which is at rest in the equilibrium state, $p_1=p-p_0$ and $T_1=T-T_0$ are the deviations of the pressure and temperature from their equilibrium values, the characteristic value of the electric field is $\mathbf{E}_1=\mathbf{E}-\gamma\nabla T_0$, and n_1 is the concentration of carriers with charge e.

In writing down these equations, we have used the fact that convection develops aperiodically, so the frequency of the resulting motions is $\omega=0$. In fact, the excitation condition is written mathematically as the condition that there be no imaginary part of the frequency, i.e., $\text{Im}(\omega)=0$, while the aperiodicity of the excitation means that there is no real part of the frequency, i.e., $\text{Re}(\omega)=0$.

In the customary Boussinesq approxmiation[1,3,6,11] the system of Eqs. (7)–(11) forms a system of linear differential equations with constant coefficients. This system can be transformed into a system of algebraic equations if a solution is sought in the form of a Fourier expansion with harmonics proportional to

$$\exp\left(ik_x\frac{x}{h}+ik_y\frac{y}{h}+ik_z\frac{z}{h}\right). \tag{12}$$

The solution also determines the dependence of the external parameters (such as the heating A) on the wave vector $\mathbf{k}$ (k_x,k_y,k_z). The longitudinal part of the wave vector $\mathbf{k}_\perp$ (k_x,k_y) is real because of the translational symmetry intrinsic to a layer geometry, if we assume that the x and y axes lie in the plane of the layer. Evidently, $k_\perp\propto h/l$ is determined by the dimension l of the cell along the layer, while k_z is determined by the homogeneous boundary conditions at the $z=0$ (''bottom'') and $z=h$ (''surface'') planes.

In this paper, we consider a layer of constant thickness and neglect any surface waves which may develop on the layer surface. Changes in the thickness, as well as the excitation of surface waves,[16,17] can be accounted for, but this is unimportant for the present problem of analyzing changes in the dimensional ratio of convection cells.

In this section we formulate boundary conditions closest to those which can be realized in an experiment. (See the experiments described in Refs. 1, 3, and 6, for example.) The bottom is the plane where the liquid comes into contact with a solid mass, so that

$$v_z=v_x=v_y=T_1=0. \tag{13}$$

The condition $v_z=0$ is a ''nonpenetration'' condition, i.e., the liquid does not permeate into the substrate. The conditions $v_z=v_y=0$ correspond to the ''attachment'' of a viscous liquid to the solid plane of the substrate. The condition $T_1=0$ means that the bottom of the liquid is at a constant temperature. This condition corresponds to the condition of ''absolutely easy'' heat transfer through this plane. The Biot number B, which characterizes heat transfer from the bottom of the liquid, is very large here $(\text{B}\to\infty)$.

The surface is the plane where the liquid comes into contact with a gas (air). Since the layer thickness is constant, a nonpenetration condition hold here, i.e.,

$$v_z=0. \tag{14}$$

At the surface, surface tension forces act, specifically, the thermocapillary forces create forces along the surface, i.e., they are balanced by the tangential components of the stress tensor,

$$\rho\nu\left(\frac{\partial v_z}{\partial x}+\frac{\partial v_x}{\partial z}\right)=-\sigma\frac{\partial T_1}{\partial x};$$

$$\rho\nu\left(\frac{\partial v_z}{\partial y}+\frac{\partial v_y}{\partial z}\right)=-\sigma\frac{\partial T_1}{\partial y}. \tag{15}$$

Note that if the upper surface were also held at a constant temperature, there would be no thermocapillary forces on this surface.

One can imagine various conditions for heat transfer from the surface,[3,6] but the closest to experiment would be thermal isolation, i.e., the absence of heat transfer from this surface, with $\text{B}=0$ and

$$\frac{\partial T_1}{\partial z}=0. \tag{16}$$

Numerical studies have also been made for $\text{Re}(\omega)\neq 0$, with different values of B, and with many other complications in the problem. The possible cases are reviewed in Ref. 5.

The system of Eqs. (7)–(11) or the corresponding system of algebraic equations can be used to express the variables in terms of one another and, subsequently, the condition for the existence of a nontrivial solution satisfying homogeneous boundary conditions makes it possible to determine k_z as a function of $k_\perp$ and $\mathcal{M}$. Knowledge of these values makes it possible to examine the condition for existence of a nontrivial solution of Eqs. (7)–(11) (this condition is $\text{Im}(\omega)=0$) as an (instability) excitation condition, i.e., a condition expressing the value of an ''external'' parameter, the heating A (and, therefore, $\mathcal{R}$, $\mathcal{M}$, and $\mathcal{E}$ for a given layer of a given liquid), as a function of l/h. Subsequently, minimization of $\partial A/\partial k_\perp=0$ gives l/h for the cell which arises at the moment of excitation.

The dependence of $w=k_\perp^2/k_z^2$ on the layer thickness h enters the calculations through the relations

$$\mathcal{R}=\left(\frac{h}{h_{RI}}\right)^3 I;\quad \mathcal{R}=\left(\frac{h}{h_{RM}}\right)^2\mathcal{M};\quad \mathcal{M}=\frac{h}{h_c}I. \tag{17}$$

Some typical results from a numerical solution are shown in Fig. 1. At the time of excitation one has $w=w^*$. In the regions $h\simeq h_c$ and $h\simeq h_{RM}$, the variation in the dimensional ratio of a cell which has formed at the time of excitation has the form typical of second-order phase transitions.[18]

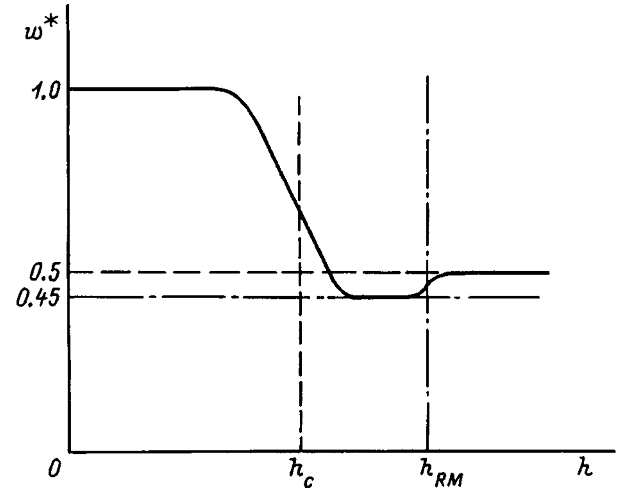

FIG. 1. The variation of w^*, the ratio of the square of the dimensions of a convection cell along and perpendicular to the layer, with the thickness h of the liquid layer.

The rearrangement of the entire convection zone when the layer thickness is changed is similar to an allotropic change in a crystal lattice.

4. EXACT SOLUTION OF THE PROBLEM FOR THE CASE OF TWO FREE BOUNDARIES HELD AT A CONSTANT TEMPERATURE

It has been shown for the mechanism induced by the buoyancy force[1] and for that induced by a thermoelectric field[9] that when the layer has two free boundary surfaces held at a constant temperature, one can avoid solving the double eigenvalue problem, and it is found that $k_z=\pi$ is the only possible value for k_z at the time of excitation.

Then the condition for existence of a nontrivial solution of Eqs. (7)–(11) is converted into the apprehensible excitation condition

$$-(1+w)^3+\left(\frac{h}{\pi h_{RI}}\right)^3\frac{I}{\pi}w+\frac{I^2}{\pi^2}(1+w)w=0 \tag{18}$$

or, equivalently, to

$$-(1+w)^3+\frac{\mathcal{R}}{\pi^4}w+\left(\frac{\pi h_{RI}}{h}\right)^6\frac{\mathcal{R}^2}{\pi^8}(1+w)w=0. \tag{19}$$

Evidently, Eq. (18) is convenient for studying excitation in layers with thicknesses $h \ll h_{RI}$, while Eq. (19) is convenient for studying the conditions for excitation of motion in layers with $h \gg h_{RI}$. Recall that, since the boundary surfaces of the layer are kept at constant temperatures $T_1=0$ for $z=0,h$, the surface tension (thermocapillarity) has no effect on the excitation conditions (see condition (15)).

Minimizing the dimensionless number I (or $\mathcal{R}$, respectively), we find

$$-3(1+w)^2+\left(\frac{h}{\pi h_{RI}}\right)^3\frac{I}{\pi}+\frac{I^2}{\pi^2}(1+2w)=0 \tag{20}$$

or

$$-3(1+w)^2+\frac{\mathcal{R}}{\pi^4}+\left(\frac{\pi h_{RI}}{h}\right)^6\frac{\mathcal{R}^6}{\pi^8}(1+2w)=0. \tag{21}$$

Eliminating I from Eqs. (18) and (20) (or $\mathcal{R}$ from Eqs. (19) and (21)), we find that w^* is the root of the equation

$$(1+w)^4(1-w)^2-\left(\frac{h}{\pi h_{RI}}\right)^6 w^2(2w-1)=0. \tag{22}$$

When h/h_{RI} is varied from 0 to ∞, w^* varies from 1 to 1/2. For $h_{RI}/h=1$, we find $w^*=0.81$ with $I^*\simeq 1.6\pi$. The exact solution for other values of h/h_{RI} shows that the dependence of w^* on h actually does reproduce the numerical solution (see Fig. 1) with, of course, h_c replaced by h_{RI} and without the segment from h_c to h_{RM}. These results also confirm that for heating from below, the excitation conditions are eased as a result of the joint action of the excitation mechanisms.[9]

5. ANALYSIS OF EXPERIMENTAL DATA

The only experiments known up to now with a sufficiently thin layer of liquid heated from below are the classical experiments of Bénard[19] (see also the detailed discussion in Ref. 1). Bénard did his experiments with a layer of spermaceti wax. A 1-mm-thick layer of wax was placed on a flat metal tray and heated by steam from below. An infinite plane layer heated from below was modeled in this way. The lower boundary was solid and isothermal. The upper boundary was the interface with air, i.e., free and thermally insluated. The calculations of Sec. 3 were done under precisely these conditions.

Data on spermaceti can be found elsewhere.[20–22] Estimates show that the thickness h of the layer in the experiment lies within the interval $h_c<h<h_{RM}$, with $h_c \simeq 10\ \mu$m (see Eq. (4)) and $h_{RM}\simeq 2$ mm (see Eq. (2)). This implies that the main mechanism leading to the formation of Bénard cells is thermocapillarity.[2]

The quantities most conveniently measured in such experiments are the size of the cell along (l) and perpendicular to (h) the layer. The success of Rayleigh's theory was explained by the fact that the l/h ratios for cells excited by the buoyancy force and thermocapillarity are similar ($\simeq$0.7 and 0.65, respectively). It was possible to identify the predominant mechanism only by analyzing the temperature difference T_h-T_c (see Ref. 1 about the history of this question).

The effect of the thermoelectric mechanism in the thinnest experimentally observable layers with thicknesses $h\simeq 10h_c\simeq 0.1$ mm can be taken into account based on the problem posed in Section 3. Under these experimental conditions, we find $\mathcal{M}\simeq\mathcal{M}(1-0.4I^2)$, while $I\simeq 0.33$. The change in the ratio l/h of the cell at the time of excitation is also small. However, the direction of the changes in the temperature difference and, especially, of the change in the size ratio of the cell induced by the thermoelectric mechanism is correct. In the thinnest layers, cells with larger l/h and smaller T_h-T_c were observed than predicted by the theory of the thermocapillary mechanism.

It is possible that the thermoelectric mechanism explains the phenomena responsible for ''aging'' of liquid crystal displays.[23] A detailed analysis of the aging, however, requires that the special properties of the liquid crystal state[24] and the design features of the device be taken into account. All of the many remaining experiments on heating from below have been done with layers whose thicknesses are such that neither thermocapillarity nor thermoelectric phenomena are of any significance.

CONCLUSION

Convection cells, their excitation, and the transition from an isotropic medium to a medium with a spatial–temporal structure constitute a classic example of a synergetic phenomenon.[25] The realignment of the spatial–temporal structure studied in this paper is yet another detail of this phenomenon. The onset of motion leads to a sudden increase in heat transfer, so that the liquid is always under conditions close to the excitation conditions at low supercriticalities.[11]

The interaction of the various excitation mechanisms or, in the language of synergetics, of the different subsystems, leads to a rearrangement of the entire convection zone owing to a change in the ratio of the dimensions of an individual cell when the layer thickness is changed. If the thickness of the heated layer is close to critical, then even a slight change in its thickness, for example, by evaporation, may cause a substantial rearrangement of the convection zone at a constant external heating.

[1] S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability*, Oxford Univ. Press (1960).

[2] J. K. A. Pearson, J. Fluid Mech. **4**, 489 (1958).

[3] V. G. Levich and V. S. Krylov, Ann. Rev. Fluid Mech. **1**, 293 (1969).

[4] I. V. Ioffe, N. V. Kalinin, and E. D. Éĭdel’man, Pis’ma Zh. Tekh. Fiz. **2**, 395 (1976) [Sov. Tech. Phys. Lett. **2**, 153 (1976)].

[5] E. D. Éĭdel’man, Usp. Fiz. Nauk **149**, 1279 (1995).

[6] G. Z. Gershuni and E. M. Zhukhovitskiĭ, *Convective Instability of Incompressible Fluids* [in Russian], Nauka, Moscow (1972).

[7] E. D. Éĭdel’man, Zh. Éksp. Teor. Fiz. **103**, 1633 (1993) [JETP **76**, 802 (1993)].

[8] D. A. Nield, J. Fluid Mech. **19**, 341 (1964).

[9] E. D. Éĭdel’man, Zh. Tekh. Fiz. **64**(4), 29 (1994) [Tech. Phys. Lett. **39**(2), 364 (1994)].

[10] E. D. Éĭdel’man, Fiz. Tekh. Poluprovodn. **28**, 1535 (1994) [Semiconductors **28**, 858 (1994)].

[11] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon Press, Oxford, 1987; Nauka, Moscow, 1988).

[12] *Physical Constants. A Handbook* [in Russian], Moscow (1990).

[13] V. I. Nezhelko and L. I. Floka, *Surface Tension of Liquid Metals and Alloys. A Handbook* [in Russian], Moscow (1991).

[14] M. Kastler, *Liquid Semiconductors* [in Russian], Mir, Moscow (1980).

[15] F. V. Bunkin, N. A. Kirichenko, B. S. Luk’yanchuk *et al.*, Kvantovaya Élektron. (Moscow) **9**, 1848 (1982) [Sov. J. Quantum Electron. **12**, 435 (1982)].

[16] E. D. Éĭdel’man, Zh. Tekh. Fiz. **64**(5), 1 (1994) [Tech. Phys. Lett. **39**(3), 455 (1994)].

[17] E. D. Éĭdel’man, Poverkhnost’, No. 3, 29 (1995).

[18] R. Brout, *Phase Transitions*, W. A. Benjamin, New York (1965).

[19] H. Bénard, Ann. Chim. Phys. **23**, 62 (1901).

[20] A. H. Warth, *The Chemistry and Technology of Waxes*, Reinhold, New York (1956).

[21] ‘‘High molecular unsaturated alcohols from sperm whale oil’’ [in Russian], Dep. in VINITI, Moscow (1966).

[22] T. Musaev, Author’s Abstract of Candidate’s Dissertation [in Russian], Grozny (1982).

[23] K. D. Vinokur, D. G. Sikharulidze, G. S. Chilaya, and Z. M. Élashvili, *Liquid Crystals with Helical Structure and their Applications in Display Units* [in Russian], Tbilisi (1988).

[24] E. D. Éĭdel’man, Fiz. Tverd. Tela **37**, 160 (1995) [Phys. Solid State **37**, 86 (1995)].

[25] H. Haken, *Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices* (Springer-Verlag, New York, 1983; Mir, Moscow, 1980).

Translated by D. H. McNeill

Calculating shock-wave processes in bubbly liquids

V. S. Surov

Chelyabinsk State University, 454021 Chelyabinsk, Russia

(Submitted November 25, 1996; resubmitted March 10, 1998)
Zh. Tekh. Fiz. **68**, 12–19 (November 1998)

A complete solution is given to the problem of the decay of an arbitrary discontinuity in a one-velocity model for a bubbly liquid and is used to analyze the propagation and interaction of shock waves in liquids with gas bubbles. *© 1998 American Institute of Physics.* [S1063-7842(98)00311-0]

INTRODUCTION

The propagation of waves in two-phase gas–liquid media with a bubbly structure is the subject of an extensive literature.[1–3] In this paper a study is made of wave phenomena in bubbly liquids using a gas dynamic approach which neglects the small scale fluctuations of the bubbles and which is justified if gas–liquid suspensions with a small amount of gas in the mixture are under consideration. Usually it is assumed that gas bubbles in a shock wave are compressed isothermally, so the Campbell–Pitcher model[4] of an equilibrium dispersion medium is customarily employed in the calculations. There are, however, some experiments for which this model yields unsatisfactory results. These experiments include those in which an ''anomalously'' low pressure level has been noticed in the shock wave reflected from a barrier when a small amount of surfactant is added to the water–bubble mixture.[5] Similar results occur when the initial mixture is diluted with glycerine, as well as with enhanced initial pressures.[6] As will be shown below, the Rakhmatulin adiabatic model[7] for a bubbly fluid is to be preferred for these experiments. A complete solution of the problem of the decay of an arbitrary discontinuity is given here for this model and the results are used to analyze a number of self similar problems of practical importance. An analysis of the results given here can be used to demonstrate the conditions for the validity of one or the other model of a bubbly liquid.

MODEL OF A BUBBLY LIQUID

Let us consider a one-velocity, single-pressure model of a continuous medium for a binary heterogeneous mixture consisting of compressible gaseous and liquid fractions. The one dimensional flow of each of the constituents of the mixture of components obeys the equations

$$\frac{\partial \alpha_g\rho_g^0}{\partial t}+\frac{\partial \alpha_g\rho_g^0 u}{\partial x}=0,\quad \frac{\partial \alpha_g\rho_g^0 u}{\partial t}+\frac{\partial(\alpha_g p+\alpha_g\rho_g^0 u^2)}{\partial x}=0,$$
$$\frac{\partial \alpha_g\rho_g^0(\varepsilon_g+0.5u^2)}{\partial t}+\frac{\partial[\alpha_g\rho_g^0 u(\varepsilon_g+0.5u^2)+\alpha_g p u]}{\partial x}=0; \tag{1}$$

$$\frac{\partial \alpha_l\rho_l^0}{\partial t}+\frac{\partial \alpha_l\rho_l^0 u}{\partial x}=0,\quad \frac{\partial \alpha_l\rho_l^0 u}{\partial t}+\frac{\partial(\alpha_l p+\alpha_l\rho_l^0 u^2)}{\partial x}=0,$$
$$\frac{\partial \alpha_l\rho_l^0(\varepsilon_l+0.5u^2)}{\partial t}+\frac{\partial[\alpha_l\rho_l^0 u(\varepsilon_l+0.5u^2)+\alpha_l p u]}{\partial x}=0, \tag{2}$$

which express the conservation of mass, momentum, and energy. Here we have used the following notation: ρ_g and ρ_l are the reduced densities, $\alpha_g=\rho_g/\rho_g^0$ and $\alpha_l=\rho_l/\rho_l^0$ are the volume fractions, ρ_g^0 and ρ_l^0 are the true densities, ε_g and ε_l are the specific internal energies (the subscripts g and l correspond to the gaseous and liquid components), and p and u are the pressure and velocity, which are common to all components of the mixture. Adding the corresponding equations of the systems (1) and (2), and noting that $\alpha_g+\alpha_l=1$, we obtain the system

$$\frac{\partial\rho}{\partial t}+\frac{\partial\rho u}{\partial x}=0,\quad \frac{\partial\rho u}{\partial t}+\frac{\partial(p+\rho u^2)}{\partial x}=0,$$
$$\frac{\partial\rho E}{\partial t}+\frac{\partial(\rho E+p)u}{\partial x}=0, \tag{3}$$

where $E=\varepsilon+0.5u^2$ is the total specific energy of the mixture, $\rho\varepsilon=\rho_g\varepsilon_g+\rho_l\varepsilon_l$ is the internal energy per unit volume of the mixture, and $\rho=\rho_g+\rho_l=\alpha_g\rho_g^0+\alpha_l\rho_l^0$ is the density of the mixture.

We shall assume that the thermodynamic parameters of the liquid obey the binomial equation of state

$$\varepsilon_l=\frac{p-c_{l*}^2(\rho_l^0-\rho_{l*})}{(\gamma_l-1)\rho_l^0}, \tag{4}$$

where γ_l, ρ_{l*}, and c_{l*} are constants which determine its properties.

In particular, for water and glycerine, they have the following values: $\gamma_l=5.59$, $\rho_{l*}=1000$ kg/m^3, $c_{l*}=1515$ m/s and $\gamma_l=7.85$, $\rho_{l*}=1260$ kg/m^3, $c_{l*}=1923$ m/s, respectively. We shall assume that the gas is ideal with an adiabatic index γ_g for which the equation of state has the form

$$\varepsilon_g=\frac{p}{(\gamma_g-1)\rho_g^0}.$$

Thus, the equation of state of the medium takes the form

$$\varepsilon=\frac{\alpha_g p}{(\gamma_g-1)\rho}+\frac{\alpha_l[p-c_{l*}^2(\rho_l^0-\rho_{l*})]}{(\gamma_l-1)\rho}. \tag{5}$$

We shall assume that the functions $\rho_g^0=\phi_g(p)$ and $\rho_l^0=\phi_l(p)$ are known. In particular, for the equation of state studied here, in the case of isentropic processes, the density of the liquid component is related to the pressure by the equation

$$\rho_l^0=\phi_l(p)=\rho_{l0}^0\left(\frac{p+p_*}{p_0+p_*}\right)^{\frac{1}{\gamma_l}},\quad p_*=\rho_{l*}c_{l*}^2/\gamma_l. \tag{6}$$

For the gaseous component we have an analogous dependence,

$$\rho_g^0=\phi_g(p)=\rho_{g0}^0\left(\frac{p}{p_0}\right)^{\frac{1}{\gamma_g}}. \tag{7}$$

The subscript 0 denotes values of the parameters for some initial state. The conservation of mass implies that

$$\frac{\rho_0}{\rho}=\frac{\alpha_{g0}\rho_{g0}^0}{\phi_g(p)}+\frac{\alpha_{l0}\rho_{l0}^0}{\phi_l(p)}. \tag{8}$$

Let us calculate the speed of sound in the mixture. After substituting Eqs. (6) and (7) in Eq. (8), we obtain the equation of state of the medium,

$$\frac{\rho_0}{\rho}=\alpha_{l0}\left(\frac{p_0+p_*}{p+p_*}\right)^{\frac{1}{\gamma_l}}+\alpha_{g0}\left(\frac{p_0}{p}\right)^{\frac{1}{\gamma_g}}. \tag{9}$$

Differentiating Eq. (9) with respect to the density ρ and noting that $c^2=dp/d\rho$, we find the square of the speed of sound in the medium to be

$$c^2=\left(\frac{\rho}{\rho^2}\right)\left[\frac{\alpha_{l0}}{\gamma_l(p+p_*)}\left(\frac{p_0+p_*}{p+p_*}\right)^{\frac{1}{\gamma_l}}+\frac{\alpha_{g0}}{\gamma_g p}\left(\frac{p_0}{p}\right)^{\frac{1}{\gamma_g}}\right]. \tag{10}$$

The corresponding Riemann invariants have the form

$$s=u-\sigma(p);\quad r=u+\sigma(p),$$

$$\sigma(p)=\int_{p_1}^{p}\frac{dp}{\rho c}=\int_{p_1}^{p}\frac{1}{\sqrt{p_0}}\left[\frac{\alpha_{l0}}{\gamma_l(p+p_*)}\left(\frac{p_0+p_*}{p+p_*}\right)^{\frac{1}{\gamma_l}}+\frac{\alpha_{g0}}{\gamma_g p}\left(\frac{p_0}{p}\right)^{\frac{1}{\gamma_g}}\right]^{1/2}dp. \tag{11}$$

We shall assume that during shock compression each component of the mixture is compressed in accordance with its own shock adiabat. For the binomial equation of state (4), the corresponding shock adiabat has the form

$$\rho_l^0=\rho_{l0}^0\frac{\chi_l(p+p_*)+p_0+p_*}{\chi_l(p_0+p_*)+p+p_*},\quad \chi_l=\frac{\gamma_l+1}{\gamma_l-1}. \tag{12}$$

For the gaseous component, we have an analogous formula,

$$\rho_g^0=\rho_{g0}^0\frac{\chi_g p+p_0}{\chi_g p_0+p},\quad \chi_g=\frac{\gamma_g+1}{\gamma_g-1}. \tag{13}$$

Thus, with Eqs. (12) and (13) the equation of the shock adiabat for the mixture takes the form

$$\frac{\rho_0}{\rho}=\alpha_{l0}\frac{\chi_l(p_0+p_*)+p+p_*}{\chi_l(p+p_*)+p_0+p_*}+\alpha_{g0}\frac{\chi_g p_0+p}{\chi_g p+p_0}. \tag{14}$$

When an isothermal variant of the Rakhmatulin model is used, where the compression of the gas in the bubbles as they cross the shock front is assumed to be isothermal, the corresponding equation of state of the mixture has the form

$$\frac{\rho_0}{\rho}=\alpha_{l0}\frac{\chi_l(p_0+p_*)+p+p_*}{\chi_l(p+p_*)+p_0+p_*}+\alpha_{g0}\frac{p_0}{p}. \tag{15}$$

Note that, as opposed to the Campbell–Pitcher approach,[4] the model used here includes the compressibility of the liquid.

In its external form Eq. (3) is the same as the corresponding gas dynamic equations for a single phase medium; thus, the methods developed for gas dynamics can be used for one-velocity gas–liquid flows. The arbitrary discontinuity decay problem or the Riemann problem play an important role here. It can be shown that a bubbly liquid with the equation of state (5) belongs to the class of normal ''gases'' which satisfy the Bethe-Bailey inequalities,[8] $\varepsilon_{VS}<0$ and $\varepsilon_{VV}<0$, where S is the entropy and $V=1/\rho$ is the specific volume. Thus, in this medium the shock waves are stable and, furthermore, the solution to the arbitrary discontinuity decay problem is unique.[9] We note also that the arbitrary discontinuity decay problem serves as a fundamental element of Godunov computational schemes.[8] Inserting the arbitrary discontinuity decay problem algorithm given in this paper in schemes of this type makes it possible to greatly extend the range of problems that can be studied and, in particular, to proceed to the study of multidimensional problems, which is problematic when other models for bubbly liquids are used. A complete solution of the arbitrary discontinuity decay problem for a bubbly liquid is given in the Appendix.

INTERACTION OF A SHOCK WAVE WITH A SOLID WALL

Let a shock wave with known pressure p_s behind its front propagate through a stationary, homogeneous gas–liquid mixture with a volume gas concentration α_{g0} and pressure p_0. The remaining parameters of the shock wave, denoted by the subscript s, are calculated from the Rankine–Hugoniot relations using the equation of state (14):

$$\rho_s=\frac{\rho_0}{\alpha_{l0}\lambda_l+\alpha_{g0}\lambda_g},\quad u_s=\sqrt{\frac{(p_s-p_0)(p_s-p_0)}{p_s p_0}},$$

$$D_s=\frac{\rho_s u_s}{\rho_s-\rho_0},\quad \alpha_{ls}=\frac{\alpha_{l0}\lambda_l}{\alpha_{g0}\lambda_g+\alpha_{l0}\lambda_l},$$

$$\lambda_l=\frac{\chi_l(p_0+p_*)+p_s+p_*}{\chi_l(p_s+p_*)+p_0+p_*},\quad \lambda_g=\frac{\chi_g p_0+p_s}{\chi_g p_s+p_0}. \tag{16}$$

Here D_s is the shock velocity and u_s is the velocity of the mixture behind the shock front. If a solid wall is placed in the path of the shock, then after it is reflected, the parameters of the mixture at the barrier, denoted by subscript r, can also be calculated using the Rankine–Hugoniot relations:

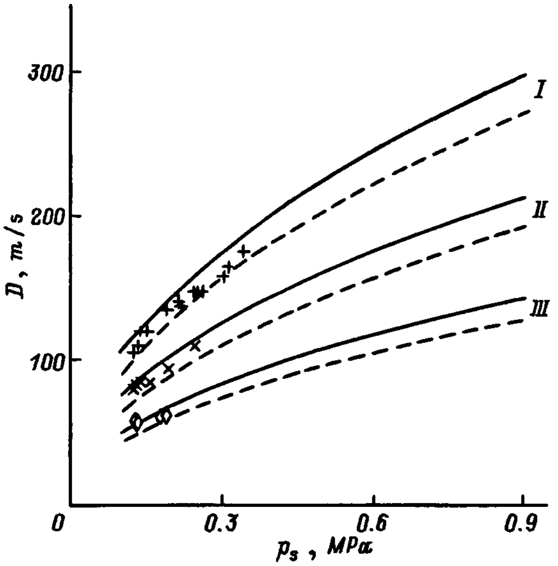

FIG. 1. $D(p_s)$ curves for a water–glycerine solution with air bubbles, calculated according to the adiabatic (smooth curves) and isothermal (dashed curves) models: $\alpha_{g0}=0.01$, $\alpha_{gl}=0.85$ (*I*); $\alpha_{g0}=0.02$, $\alpha_{gl}=0.85$ (*II*); $\alpha_{g0}=0.05$, $\alpha_{gl}=0.5$ (*III*); symbols correspond to experiment.[10]

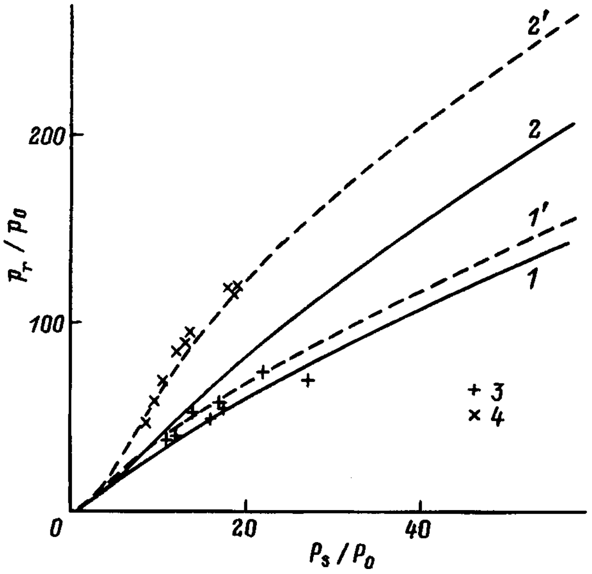

FIG. 2. p_r/p_0 as a function of p_s/p_0 from experiment[6] (*3,4*) and as calculated using the adiabatic (*1,2*) and isothermal (*1′,2′*) models for $\alpha_{g0}=0.005$ (*1,1′,3*) and 0.03 (*2,2′,4*).

$$(\rho_s-\rho_r)D_r=\rho_s u_s, \quad \rho_s u_s D_r=p_s+\rho_s u_s^2-p_r,$$

$$\frac{\rho_s}{\rho_r}=\frac{\alpha_{gs}(\chi_g p_s+p_r)}{\chi_g p_r+p_s}+\frac{\alpha_{ls}(\chi_l(p_s+p_*)+p_r+p_*)}{\chi_l(p_r+p_*)+p_s+p_*}. \quad (17)$$

After D_r and ρ_r are eliminated, the system of Eqs. (17) reduces to a cubic equation in p_r, which was solved numerically.

Let us compare the calculations with experiment. Figure 1 shows the calculated and experimental[10] dependences of the shock velocity in a water–glycerine solution with air bubbles on the pressure behind the shock front for different gas concentrations in the mixture ($p_0=0.1$ MPa). The suspension was assumed to have three components, so the calculations were done by modifying Eq. (16) to account for the presence of a third component, glycerine, in the mixture. In particular, for the adiabatic model the density of the mixture behind the shock front was calculated using the formulas

$$\rho_s=\frac{\rho_0}{\alpha_{l0}\lambda_l+\alpha'_{l0}\lambda'_l+\alpha_{g0}\lambda_g},$$

$$\lambda'_l=\frac{\chi'_l(p_0+p'_*)+p_s+p'_*}{\chi'_l(p_s+p'_*)+p_0+p'_*},$$

where the primes denote the parameters of glycerine.

Figure 2 compares the pressure jump during reflection of a shock from the wall in a water–nitrogen suspension ($\gamma_g=1.4$; $\rho^0_{g0}=1.15$ kg/m^3) as a function of the shock intensity for volume concentrations $\alpha_{g0}=0.005$ and 0.03 of gas in the mixture ($p_0=0.1$ MPa), calculated using the adiabatic and isothermal models and obtained by experiment.[6] For $\alpha_{g0}=0.005$ the curves calculated using these models are close to one another and agree with experiment. The adiabatic model calculations with $\alpha_{g0}=0.03$ give lower values of the reflection coefficient than observed experimentally or calculated with the isothermal model.

Figure 3 shows plots of the reflection coefficients p_r/p_s as functions of the incident shock intensity p_s/p_0 calculated using the adiabatic and isothermal models for a water–air suspension with a gas concentration $\alpha_{g0}=0.05$ in the mixture for $p_0=0.1$ and 1.0 MPa. Experimental data from Ref. 6 are also shown there. Note the closeness of the experimental data to the adiabatic model calculations for the higher initial pressure $p_0=1.0$ MPa. For $p_0=0.1$ MPa the reflection coefficient for $p_s/p_0<10$ is the same as that calculated by the isothermal model, but when the incident shock intensity is

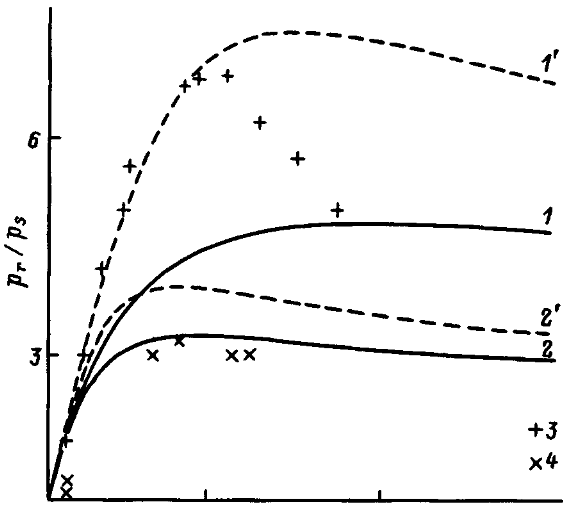

FIG. 3. p_r/p_s as a function of p_s/p_0 from experiment[6] (*3,4*) and as calculated for $\alpha_{g0}=0.05$ using the adiabatic (*1,2*) and isothermal (*1′,2′*) models for initial pressures $p_0=0.1$ (*1,1′,3*) and 1.0 MPa (*2,2′,4*).

increased further, the adiabatic model gives better results (Fig. 3).

These data can be explained as follows: in the experiments,[6] two-phase mixtures with rather large air bubbles having diameters of 3-4 mm were used. It is known[11] that air bubbles larger than 2 mm are destroyed in shocks with intensities $p_s/p_0>5$. Breakup of the bubbles is accompanied by faster heat transfer between the gas and liquid owing both to an increase in the contact surface area and to more intense mixing of the gas within the bubbles that are breaking up. For this case, the isothermal model gives results that are close to experiment. If we exclude the possibility of bubble breakup, then heat transfer is more difficult and the pressure level observed in the experiments ends up close to that calculated using the adiabatic model of a dispersion medium. It is known that when the initial pressure is increased, the resistance to breakup on the part of the bubbles is greater, even for the quite large bubbles;[6] thus, as p_0 is raised, the pressure level in the shock wave should approach that of the adiabatic model calculations. This tendency is confirmed by the data of Fig. 3. An analogous effect can be obtained by adding a surfactant to the initial gas–liquid mixture[5] or diluting a water–air suspension with glycerine.[6] Another stabilizing factor which impedes breakup of the bubbles is replacing the air in the bubbles with a lighter gas (such as hydrogen).[12] In all these cases it is recommended that the adiabatic model be used instead of the isothermal model. As for the data of Fig. 3 at an initial pressure $p_0=0.1$ MPa, the results of this type evidently depend on the experimental conditions. In fact, the experiments in Ref. 6 were conducted on a shock tube with a short measurement length. High-amplitude incident shocks have a high velocity, so the time the sensor lies within the region behind the shock front is short. Because of the finite rate of interphase heat exchange, the gas cannot reach the temperature of the liquid, so the reflection coefficient is lower than that obtained by the isothermal model. For this reason, as the amplitude of the incident shock is increased, the reflection coefficient approaches the value calculated with the adiabatic theory. Similar results apply to oblique, as well as normal, reflection of shocks from a boundary.[13]

INTERACTION OF A SHOCK WAVE WITH A BUBBLE SCREEN LOCATED NEAR A BARRIER

The effect of a short shock wave incident from a pure (without bubbles) liquid on a bubble screen, used, for example, as protection against explosive loading, has been discussed elsewhere.[12] We shall examine the problem of a long shock wave having a profile in the form of a semi-infinite step and propagating through a gas–liquid mixture with a gas content α_{g0} and incident on a screen of thickness L_0 sited at a barrier. We shall assume that the volume concentration $\alpha_{g0}^{\mathrm{sc}}$ of the gas in the screen obeys $\alpha_{g0}^{\mathrm{sc}}>\alpha_{g0}$.

Figure 4 is a schematic illustration of the shock-wave picture resulting from this interaction. At time $t=0$ the incident shock reaches the screen, causing formation of a rarefaction wave (RW) which is reflected from the layer and a shock SW_1 which moves along the screen at velocity D_1.

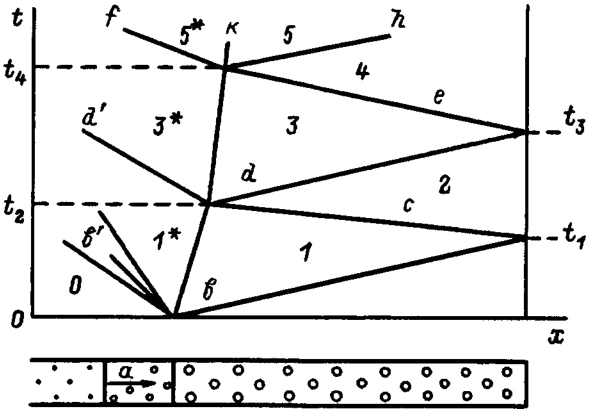

FIG. 4. $x-t$ diagram for the interaction of a shock wave with a bubble screen: (a) shock wave SW, (b) SW_1, (b′) rarefaction wave RW, (c) SW_2, (d) SW_3, (d′) SW_3^*, (e) SW_4, (f) SW_5^*, (h) SW_6.

The parameters of these waves were calculated by solving the problem of the decay of an arbitrary discontinuity. At the time $t_1=L_0/D_1$, SW_1 reaches the barrier surface and is reflected from it, forming a shock SW_2, which, at the time

$$t_2=\frac{L_0+t_1|D_2|}{u_1+|D_2|}$$

is, in turn, reflected from the contact boundary of the bubble screen. Subsequent reflections, also calculated using the corresponding problem of the decay of an arbitrary discontinuity, occur at times

$$t_3=\frac{L_0+t_2(D_3-u_1)}{D_3},\quad t_4=\frac{L_0-t_2u_1+t_3|D_4|}{|D_4|},$$

$$t_5=\frac{L_0-t_2u_1+t_4D_5}{D_5},\quad t_6=\frac{L_0-t_2u_1+t_4u_3+t_5|D_6|}{u_3+|D_6|},$$

$$t_7=\frac{L_0-t_2u_1+t_4u_3+t_6(D_7-u_3)}{D_7},\ \ldots .$$

Here D_k and u_k are the velocities of the kth reflected shock and contact discontinuity (CD) after reflection of SW_{k-1} from the free surface of the bubbly layer.

Figure 5 shows the pressure at the barrier as a function of time for an incident shock with $p_s=1.0$ MPa ($p_0=0.1$ MPa), calculated with the models used in this paper for different gas concentrations in the screen ($L_0=1$ mm) and outside it. Figure 6 shows the time variations in the volume fraction of gas at the wall, as well as the relative thickness of the screen for one version of the calculations.

An analysis of the data in Figs. 5 and 6 shows that the screen is compressed over time and the volume fraction of gas in it decreases. Here the degree of compression of the screen is higher according to the isothermal model. The pressure at the barrier calculated by the adiabatic Rakhmatulin model, as in Ref. 2, approaches the value in the reflected shock without a screen. According to the isothermal model, the maximum pressure is higher than that calculated in the adiabatic model. With increasing p_s the delay time provided by the bubble screen is reduced. Thus, for a shock with a post-shock pressure $p_s=5$ MPa propagating through a liquid

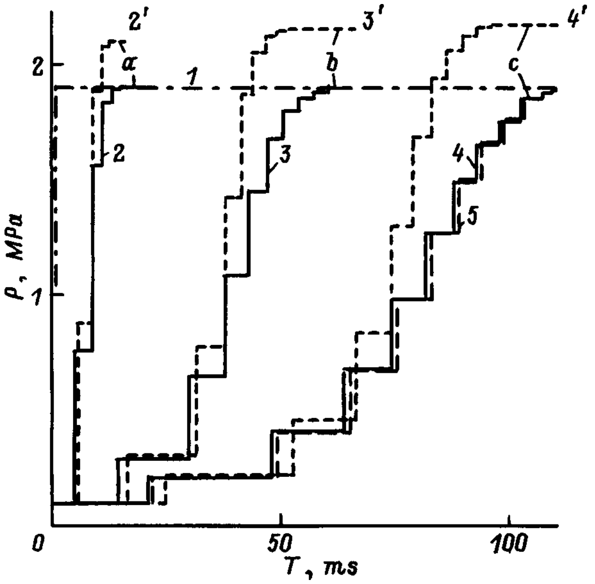

FIG. 5. $p(t)$ at the barrier according to the adiabatic (*2–4*) and isothermal (*2′–4′*) models for (a) $\alpha_{g0}=0$, $\alpha_{g0}^{\rm sc}=0.01$ (*2,2′*); (b) $\alpha_{g0}=0.01$, $\alpha_{g0}^{\rm sc}=0.1$ (*3,3′*); (c) $\alpha_{g0}=0$, $\alpha_{g0}^{\rm sc}=0.1$ (*4,4′*); (1) $p(t)$ without a screen; (*5*) $p(t)$ for a hydrogen screen ($\alpha_{g0}=0$, $\alpha_{g0}^{\rm sc}=0.1$).

without bubbles ($\alpha_{g0}=0$, $p_0=0.1$ MPa), and interacting with a bubble screen having a thickness of 1 m and volume gas contents $\alpha_{g0}^{\rm sc}=0.01$ and 0.1, the time delays (until the pressure reaches its maximum value) are 3 and 18 ms. The computational results depend only weakly on the type of gas used to form the bubbles in the screen. This can be seen in Fig. 5, which shows adiabatic model calculations of the pressure at the barrier in a mixture of a liquid with hydrogen bubbles. For hydrogen the adiabatic index and initial density were taken to be $\gamma_g=1.33$ and $\rho_{g0}^0=0.0148$ kg/m^3. Note that, depending on the heat transfer conditions between the gas and liquid, it will be necessary to use one or the other model for the bubbly liquid.

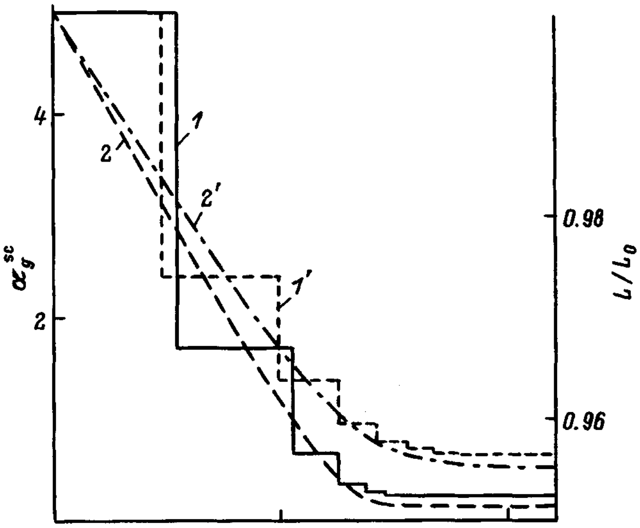

FIG. 6. $\alpha_{g0}^{\rm sc}(t)$ (*1,1′*) at a wall and $L/L_0(t)$ (*2,2′*) calculated using the adiabatic (*1,2*) and isothermal (*1′,2′*) models.

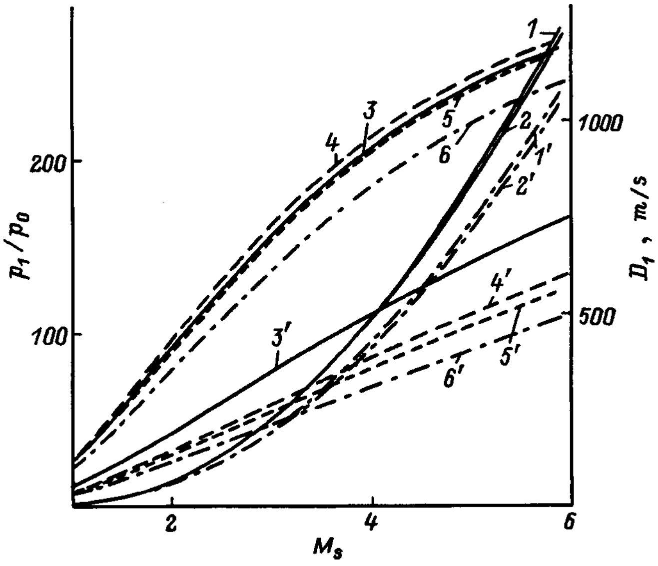

FIG. 7. p_1/p_0 (*1,6*) and D_1 (*2–5*) as functions of M_s for $\alpha_{g0}=0.01$ for water with nitrogen (*1,2*), helium (*4*), and hydrogen (*5*) bubbles, calculated using the adiabatic model; (*2,6*) using the isothermal model; (*1′–6′*) same curves, but for $\alpha_{g0}=0.1$.

INTERACTION OF AN AIR SHOCK WAVE WITH A LAYER OF BUBBLY LIQUID

Let a plane air shock wave with a semi-infinite step profile be incident normally on a layer of uniform bubbly liquid with gas content $\alpha_{g0}^{\rm sc}$ located at a solid wall. Nitrogen, hydrogen, and helium were considered as gases filling the bubbles. For the latter, $\gamma_g=1.67$ and $\rho_{g0}^0=0.164$ kg/m^3. It is necessary to calculate the flow resulting from the interaction.

The parameters of the air behind the incident shock front, denoted by subscript s, are related to the initial parameters before the jump (subscript 0) by the Rankine–Hugoniot relations:

$$p_s=p_0\left(1+\frac{2\gamma(M_s^2-1)}{(\gamma+1)}\right),\qquad u_s=\frac{2c_{g0}(M_s^2-1)}{(\gamma+1)M_s},$$

$$\rho_s=\rho_{g0}\left(\frac{(\gamma+1)M_s^2}{2+(\gamma-1)M_s^2}\right),$$

where c_{g0}, γ and $M_s=D/c_{g0}$ are, respectively, the sound speed and adiabatic index of air and the Mach number (D is the speed of the shock front).

When the air shock reaches the contact boundary of the layer, a reflected shock SW_1^* develops and propagates through the gas away from the layer, as does a penetrating shock SW_1, which moves at speed D_1 through the bubbly liquid. The parameters of these shocks were calculated by solving the problem of the decay of an arbitrary discontinuity.

Figure 7 shows adiabatic and isothermal model calculations of the pressure jump p_1/p_0 behind the front of a shock SW_1 penetrating into a liquid with nitrogen bubbles as a function of the Mach number of the incident shock for dif-

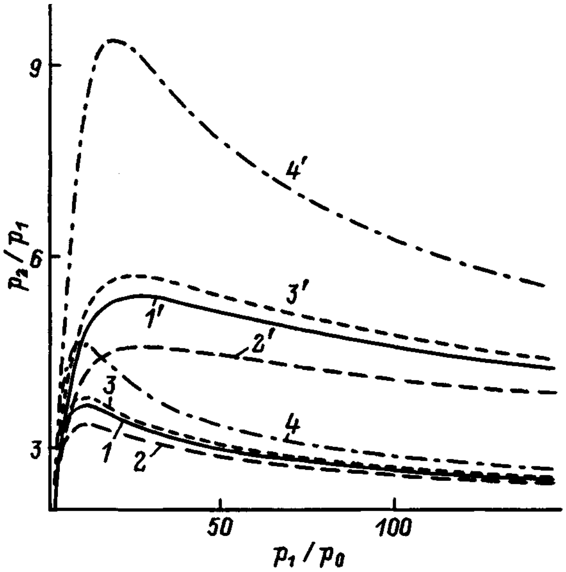

FIG. 8. p_2/p_1 as a function of p_1/p_0 for water with bubbles of nitrogen, helium and hydrogen according to the adiabatic (*1–3*) and isothermal (*4*) models for $\alpha_{g0}=0.01$; (*1′–4′*) for $\alpha_{g0}=0.1$.

ferent gas concentrations in the mixture ($p_0=0.1$ MPa). For a liquid with hydrogen and helium bubbles, the corresponding pressure curves coincided with those for nitrogen bubbles to within the accuracy of the graph. Also shown in this figure are the variations in the propagation speed of SW_1. Note that the curves calculated using the isothermal model are independent of the type of gas in the bubbles.

When SW_1 reaches the barrier, a reflected shock SW_2 is formed. Figure 8 shows the reflection coefficient p_2/p_1 as a function of the intensity p_1/p_0 of the penetrating shock SW_1, calculated using the adiabatic and isothermal models of Rakhmatulin for water with bubbles of nitrogen, helium, and hydrogen with different volume concentrations of the gas in the mixture. As can be seen from Fig. 8, the pressure in the shock reflected from the barrier according to the adiabatic model depends substantially on the type of gas filling the bubbles, while the pressure difference behind the reflected shock is greater for higher gas concentrations in the liquid. By contrast, the shape of the curves calculated using the isothermal model does not depend on the type of gas in the bubbles.

CONCLUSION

In this paper it has been shown that when modeling shock wave phenomena in bubbly liquids under conditions of developed heat transfer between the gaseous and liquid fractions, which favors the breakup of bubbles in the shock, it is necessary to use an isothermal model for the dispersion medium. In the case of inhibited heat transfer between the fractions, however, better results are obtained with an adiabatic model. The latter is also appropriate for modeling the action of short, high-power pressure pulses on bubbly liquids. For very low gas concentrations in the liquid, heat transfer processes have little effect on the shock parameters, so the results of calculations employing both models are similar.

APPENDIX A:

Let two infinite masses of bubbly liquids with volume gas concentrations $\alpha_{(1)g0}$ and $\alpha_{(2)g0}$, respectively, meet at the $x=0$ plane at the initial time ($t=0$). The pressure, velocity, and density in these media are constant and equal to $p_{(1)0}$, $u_{(1)0}$, $\rho_{(1)0}$ and $p_{(2)0}$, $u_{(2)0}$, $\rho_{(2)0}$, respectively. Without loss of generality, we shall assume that $p_{(1)0}\geqslant p_{(2)0}$. It is necessary to calculate the flow which develops for $t>0$.

It is known that if an arbitrary discontinuity is not a contact discontinuity or a shock wave, then it decays, forming either two shock waves, or a rarefaction wave and a shock wave, or two rarefaction waves.[8] Let us look at these cases in more detail.

Two shock waves. The system of equations for the shock wave on the right, which express the conservation of mass and momentum on passing through the shock front, together with the equation of state (14), has the form

$$(\rho_{(2)+}-\rho_{(2)0})D_{(2)+}=\rho_{(2)+}u_{(2)+}-\rho_{(2)0}u_{(2)0},$$

$$(\rho_{(2)+}u_{(2)+}-\rho_{(2)0}u_{(2)0})D_{(2)+}$$
$$=p_{(2)+}+\rho_{(2)+}u_{(2)+}^2-p_{(2)0}-\rho_{(2)0}u_{(2)0}^2,$$

$$\frac{\rho_{(2)0}}{\rho_{(2)+}}=\alpha_{(2)g0}\frac{\chi_{(2)g}p_{(2)0}+p_{(2)+}}{\chi_{(2)g}p_{(2)+}+p_{(2)0}}$$
$$+\alpha_{(2)l0}\frac{\chi_{(2)l}(p_{(2)0}+p_{*(2)})+p_{(2)+}+p_{*(2)}}{\chi_{(2)l}(p_{(2)+}+p_{*(2)})+p_{(2)0}+p_{*(2)}}. \tag{A1}$$

For the shock on the left we have an analogous system of equations,

$$(\rho_{(1)-}-\rho_{(1)0})D_{(1)-}=\rho_{(1)-}u_{(1)-}-\rho_{(1)0}u_{(1)0},$$

$$(\rho_{(1)-}u_{(1)-}-\rho_{(1)0}u_{(1)0})D_{(1)-}$$
$$=p_{(1)-}+\rho_{(1)-}u_{(1)-}^2-p_{(1)0}-\rho_{(1)0}u_{(1)0}^2,$$

$$\frac{\rho_{(1)0}}{\rho_{(1)-}}=\alpha_{(1)g0}\frac{\chi_{(1)g}p_{(1)0}+p_{(1)-}}{\chi_{(1)g}p_{(1)-}+p_{(1)0}}$$
$$+\alpha_{(1)l0}\frac{\chi_{(1)l}(p_{(1)0}-p_{*(1)})+p_{(1)-}+p_{*(1)}}{\chi_{(1)l}(p_{(1)-}+p_{*(1)})+p_{(1)0}+p_{*(1)}}. \tag{A2}$$

In Eqs. (A1) and (A2) the subscripts ‘‘+’’ and ‘‘−’’ denote values of the parameters of the mixture in the region of the contact discontinuity at its right and left, respectively. At the contact surface the pressure and velocity do not change, so that the following matching conditions hold:

$$u_{(2)+}=u_{(1)-}=U, \quad p_{(2)+}=p_{(1)-}=P. \tag{A3}$$

After transformation of Eqs. (A1)–(A3), we obtain an equation for calculating P,

$$u_{(1)0}-u_{(2)0}=f_{(1)1}(P)+f_{(2)1}(P), \tag{A4}$$

where

$$\begin{aligned} f_{(1)1}(P)=\Bigg[\frac{(P-p_{(1)0})}{\rho_{(1)0}}\Bigg(1-\frac{\alpha_{(1)g0}(\chi_{(1)g}p_{(1)0}+P)}{\chi_{(1)g}P+p_{(1)0}} \\ \frac{\alpha_{(1)l0}(\chi_{(1)l}(p_{(1)0}+p_{*(1)})+P+p_{*(1)})}{\chi_{(1)l}(P+p_{*(1)})+p_{(1)0}+p_{*(1)}}\Bigg)\Bigg]^{1/2}, \end{aligned}$$

$$\begin{aligned} f_{(2)1}(P)=\Bigg[\frac{(P-p_{(2)0})}{\rho_{(2)0}}\Bigg(1-\frac{\alpha_{(2)g0}(\chi_{(2)g}p_{(2)0}+P)}{\chi_{(2)g}P+p_{(2)0}} \\ -\frac{\alpha_{(2)l0}(\chi_{(2)l}(p_{(2)0}+p_{*(2)})+P+p_{*(2)})}{\chi_{(2)l}(P+p_{*(2)})+p_{(2)0}+p_{*(2)}}\Bigg)\Bigg]^{1/2}. \end{aligned} \tag{A5}$$

The desired root of Eq. (A4) was calculated numerically using a standard procedure for solving nonlinear equations. A configuration with two shock waves occurs if $u_{(1)0}-u_{(2)0}\geqslant U_*$, where $U_*=f_{(2)1}(p_{(1)0})$.

For the isothermal model, in place of Eq. (A5) one must use the equations

$$\begin{aligned} f_{(1)1}(P)=\Bigg[\frac{(P-p_{(1)0})}{\rho_{(1)0}}\Bigg(1-\frac{\alpha_{(1)g0}p_{(1)0}}{P} \\ -\frac{\alpha_{(1)l0}(\chi_{(1)l}(p_{(1)0}+p_{*(1)})+P+p_{*(1)})}{\chi_{(1)l}(P+p_{*(l)})+p_{(1)0}+p_{*(1)}}\Bigg)\Bigg]^{1/2}, \end{aligned}$$

$$\begin{aligned} f_{(2)1}(P)=\Bigg[\frac{(P-p_{(2)0})}{\rho_{(2)0}}\Bigg(1-\frac{\alpha_{(2)g0}p_{(2)0}}{P} \\ -\frac{\alpha_{(2)l0}(\chi_{(2)l}(p_{(2)0}+p_{*(2)})+P+p_{*(2)})}{\chi_{(2)l}(P+p_{*(2)})+p_{(2)0}+p_{*(2)}}\Bigg)\Bigg]^{1/2}. \end{aligned} \tag{A6}$$

Rarefaction wave and shock wave. If $u_{(1)0}-u_{(2)0}<U_*$, then the flow consists of a rarefaction wave propagating leftward from the contact surface and a shock wave moving to the right. Equations (A1) hold for the latter. Given that the Riemann *r*-invariant is constant in a rarefaction wave, we have

$$u_{(1)0}-u_{(1)-}=\int_{p_{(1)0}}^{p_{(1)-}}f_{(1)2}(p)dp, \tag{A7}$$

where

$$\begin{aligned} f_{(1)2}(p)=\frac{1}{\sqrt{\rho_{(1)0}}}\Bigg[\frac{\alpha_{(1)l0}}{\gamma_{(1)l}(p+p_{*(1)})}\bigg(\frac{p_{(1)0}+p_{*(1)}}{p+p_{*(1)}}\bigg)^{\frac{1}{\gamma_{(1)l}}} \\ +\frac{\alpha_{(1)g0}}{\gamma_{(1)g}p}\bigg(\frac{p_{(1)0}}{p}\bigg)^{\frac{1}{\gamma_{(1)g}}}\Bigg]^{1/2}. \end{aligned}$$

After a number of transformations, Eqs. (A1), (A3), and (A7) yield an equation for calculating the pressure P at the contact surface,

$$u_{(1)0}-u_{(2)0}=\int_{p_{(1)0}}^{p}f_{(1)2}(p)dp+f_{(2)1}(P), \tag{A8}$$

which, as in the case of a flow with two shock waves, is solved numerically. This configuration occurs in the case where

$$U_{**}\leqslant u_{(1)0}-u_{(2)0}<U_*,$$

where

$$U_{**}=\int_{p_{(1)0}}^{p_{(2)0}}f_{(1)2}(p)dp.$$

Two rarefaction waves. If $u_{(1)0}-u_{(2)0}<U_{**}$, then a flow with two rarefaction waves develops. Equation (A7) is valid for the rarefaction wave on the left. For the rarefaction wave on the right, it is necessary to use the conservation condition for the Riemann *s*-invariant,

$$u_{(2)0}-u_{(2)+}=\int_{p_{(2)+}}^{p_{(2)0}}f_{(2)2}(p)dp, \tag{A9}$$

where

$$\begin{aligned} f_{(2)2}(p)=\frac{1}{\sqrt{\rho_{(2)0}}}\Bigg[\frac{\alpha_{(2)l0}}{\gamma_{(2)l}(p+p_{*(2)})}\bigg(\frac{p_{(2)0}+p_{*(2)}}{p+p_{*(2)}}\bigg)^{\frac{1}{\gamma_{(2)l}}} \\ +\frac{\alpha_{(2)g0}}{\gamma_{(2)g}p}\bigg(\frac{p_{(2)0}}{p}\bigg)^{\frac{1}{\gamma_{(2)g}}}\Bigg]^{1/2}. \end{aligned}$$

After transforming Eqs. (A7) and (A9), and using Eq. (A3), the equation for calculating P takes the form

$$u_{(1)0}-\int_{p_{(1)0}}^{P}f_{(1)2}(p)dp=u_{(2)0}+\int_{p_{(2)0}}^{P}f_{(2)2}(p)dp.$$

If we set $\alpha_{(1)l0}=\alpha_{(2)l0}=0$ in the above formulas, then we obtain the classical problem of the decay of an arbitrary discontinuity in an ideal gas. Then the integrals in the governing equations redue to quadratures, and the resulting expressions are exactly the same as in Ref. 9. There is yet another class of problems in which the integrals can be reduced to quadratures, and this is the problem of the decay of an arbitrary discontinuity in a gas–liquid medium consisting of a binary mixture of an ideal gas and an incompressible liquid. This variant has been studied in detail elsewhere.[14] In other cases, the integrals are calculated numerically.

[1] A. A. Gubaĭdullin, A. I. Ivandaev, R. I. Nigmatulin *et al.*, Itogi Nauki Tekhn. Ser. Mekh. Zhidk. Gaz. (VINITI, Moscow) **17** (1982), pp. 160–259.

[2] R. I. Nigmatulin, *Dynamics of Polyphase Media* [in Russian], Part 2, Nauka, Moscow (1987), 360 pp.

[3] V. E. Nakoryakov, B. G. Pokusaev, and I. R. Shreĭber, *Wave Dynamics of Gas- and Vapor-liquid Media* [in Russian], Énergoatomizdat, Moscow (1990), 248 pp.

[4] I. J. Campbell and A. S. Pitcher, Proc. R. Soc. London Ser. A **243**, 534 (1958).

[5] B. E. Gel'fand, S. A. Gubin, S. M. Kogarko *et al.*, Inzh.-Fiz. Zh. **31**, 1080 (1976).

[6] B. E. Gel'fand, S. A. Gubin, and E. I. Timofeev, Izv. Akad. Nauk SSSR,

Mekh. Zhidk. Gaza No. 2, 174 (1978).

[7] Kh. A. Rakhmatulin, Prikl. Mat. Mekh. **33**, 598 (1969).

[8] S. K. Godunov (Ed.), *Numerical Solution of Multidimensional Problems in Gas Dynamics* [in Russian], Nauka, Moscow (1976), 400 pp.

[9] B. L. Rozhdestvenskiĭ and N. N. Yanenko, *Systems of Quasilinear Equations* [in Russian], Nauka, Moscow (1978), 688 pp.

[10] V. E. Nakoryakov, B. G. Pokusaev, I. R. Shreĭber *et al.*, in *Wave Processes in Two-Phase Systems* [in Russian], Novosibirsk (1975), pp. 54–97.

[11] B. E. Gel'fand, S. A. Gubin, S. M. Kogarko *et al.*, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza No. 4, 53 (1975).

[12] B. E. Gel'fand, S. A. Gubin, R. I. Nigmatulin *et al.*, Dokl. Akad. Nauk SSSR **235**, 292 (1977) [Sov. Phys. Dokl. **22**, 357 (1977)].

[13] R. B. Eddington, AIAA J. **8**, 65 (1970).

[14] V. S. Surov, Teplofiz. Vys. Temp. **34**, 285 (1996).

Translated by D. H. McNeill

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Ozone production in supersonic nozzles

O. A. Gordeev

Moscow Radiotechnical Institute, Russian Academy of Sciences, 113519 Moscow, Russia

V. N. Makarov, V. A. Pavlov, and O. P. Shatalov

Institute of Mechanics, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted March 31, 1997)
Zh. Tekh. Fiz. **68**, 20–26 (November 1998)

Ozone production during the efflux of oxygen from a supersonic nozzle is studied theoretically and experimentally. The formation kinetics of atomic oxygen in an electrical discharge in the nozzle is analyzed. An experiment is set up using an optimized nozzle with an electrical discharge in its supersonic section. It is shown that the highest ozone content at the nozzle output is attained when the excess concentration of oxygen atoms is produced in the supersonic section of the nozzle. © *1998 American Institute of Physics.* [S1063-7842(98)00411-5]

INTRODUCTION

The standard method for generating ozone consists of using continuous-flow discharge systems (ozonators) in which ozone is produced by the recombination of oxygen that has been partially dissociated in the discharge.[1] Despite the advantages of this method (simplicity, economy), questions still remain about increasing the yield of O_3 molecules per unit mass of feed gas (oxygen, air, mixtures of oxygen and inert additives) and raising the productivity of the ozonator. Appropriate gas dynamic control of the flow parameters using profiled nozzles opens up additional possibilities. In particular,[2] it has been shown that in the cooling of partially dissociated oxygen at temperatures of the order of 1000 K in a supersonic nozzle, ozone is produced at concentrations many orders of magnitude greater than the equilibrium value. The values estimated[2] for the absolute concentration of ozone, however, were below 10^{13} cm^{-3}, or much lower than those obtained under ordinary conditions. The problem of optimizing the supersonic part of the nozzle and the parameters of the gas in front of it so as to obtain maximum ozone yield in the equilibrium heating of oxygen mixed with an inert additive before entering the nozzle has been posed and solved previously.[3] Although the ozone concentrations were estimated to be as high as 10^{15}–10^{16} cm^{-3}, it became clear at the same time that it would be difficult to obtain a sufficiently high ozone yield with this scheme.[3] At low initial temperatures of the gas mixture the initial equilibrium concentration of O atoms and, hence, the relative concentration of the O_3 product molecules are low, while at high temperatures it is difficult to generate ozone.

A situation in which an excess concentration of atomic oxygen is created at the nozzle inlet or in its supersonic section (for example, by an electrical discharge) has been examined.[4,5] It was shown that the greatest effect is achieved when the excess oxygen concentration is created in the supersonic section of the nozzle.

In this paper, this approach is developed further by analyzing a wider range of initial temperatures in front of the nozzle inlet (lowering the gas temperature in front of the nozzle to room temperature) and examining the kinetics of atomic oxygen production in an electrical discharge in a nozzle, as well as by setting up an experiment using an optimum nozzle profile and creating an electrical discharge in its supersonic section. The task of this paper is to search for the optimum nozzle characteristics, gas parameters, and, in particular, the optimum O atom concentrations which will ensure the highest yield of ozone, and to create an experimental model of such an ozonator based on this search.

OPTIMIZING THE PARAMETERS OF A GAS DYNAMIC OZONATOR

The chemical reaction model used in this work has been described in detail elsewhere.[4] Also shown there is a complete set of reaction rate constants with a justification for choosing these reactions.

In solving the optimization problem, we examined a class of plane, wedge-shaped nozzles with (different) opening half angles Θ, throat heigts h_*, and lengths L. It was assumed that after expansion in such a nozzle, the gas flow enters a plane-parallel channel of length L_1 and that at the junction of the nozzle with the plane channel the profile was rounded off to avoid shock wave formation in the plane channel. We optimized the system to obtain a maximum molar concentration of ozone, ξ_3. If a fixed set of values of the initial pressure p_0 is used, then the optimization parameters will be ξ_1, ξ_1', ξ_2, T_0, α, L, L_1, and l', where ξ_1 and ξ_1' are the molar fractions of O atoms in front of the nozzle inlet and immediately after the concentration jump at the supersonic section of the nozzle, ξ_2 is the concentration of O_2 molecules in front of the nozzle inlet, T_0 is the initial gas temperature, $\alpha = 2\tan\Theta/h_*$, and l' is the distance from the geometrical throat to the jump in the concentrations of O and O_2. At the jump the conditions $\gamma_1 + \gamma_2 = \gamma_1' + \gamma_2'$ are satis-

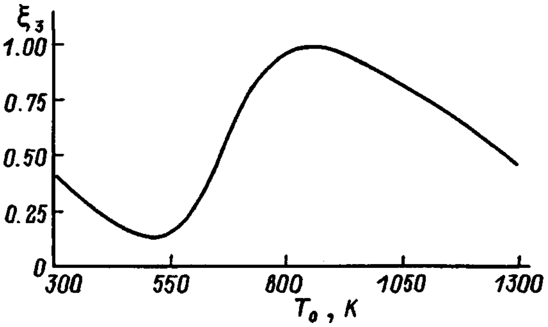

FIG. 1. Ozone concentration as a function of initial gas pressure.

fied, where $\gamma_1' > \gamma_1$. Here γ_1, γ_2, γ_1', and γ_2' are the mass fractions of oxygen atoms and molecules in front of (without a prime) and after (with a prime) the jump. The subsonic section was not optimized and was treated as a fixed profile with a radius of curvature $R=2h_*$. The optimum was sought using the method of configurations.[6]

It has been shown[4] that the main effect on ozone production is from the jump in the concentration of O atoms in the supersonic section of the nozzle. For $p_0=10$ atm and $T_0=827$ K, and with a nozzle profile having $\alpha=4.64$ cm^{-1}, $l'=0.97$ cm, and $\xi_1'=7.86\%$, the optimum molar fraction of ozone was $\xi_3=1.905\%$.

Note that the optimum conditions of Ref. 4 correspond to injecting a rather hot gas at a temperature of 827 K into the nozzle. However, by using an optimization technique that offers a combination of random and deterministic search variants we were able to observe a second maximum in the gas dynamic ozone generator. The existence of a second maximum is evident in Fig. 1. Here the ozone concentration (relative units) is plotted as a function of the initial temperature T_0 from 300 to 1300 K, while the remaining parameters correspond to the above conditions with $p_0=10$ atm and $T_0=827$ K. In Fig. 1 a concentration of unity is taken to be the ozone concentration under these conditions ($\xi_3=1.905\%$). As T_0 is reduced from the optimum ($T_0=827$ K), the ozone concentration falls off substantially because of heating of the flow owing to recombination reactions, while the length L_1 of the plane-parallel segment and the other parameters will no longer be optimal for $T_0<827$ K. At temperatures $T_0<500$ K, however, the ozone concentration again begins to rise and this continues until temperatures $T_0 \cong 300$ K. Thus, the nonmonotonicity of the temperature dependence in the presence of a jump in the atomic O concentration leads to an additional peak in the concentration of ozone molecules for low initial temperatures T_0. In this regard, a search was undertaken to find the most favorable conditions for ozone molecule formation in a supersonic nozzle at low initial temperatures $T_0=300$ K and pressures $p_0=2$ atm. In addition, the statement of the problem was simplified through the assumption that there is one jump in the atomic O concentration, which occurs only in the supersonic section of the nozzle ($\xi_1=0$), and that there is no plane-parallel section of the nozzle (i.e., $L_1=0$). Therefore, the search for an optimum was carried out for a wedge-shaped nozzle with parameters α, L, l', and ξ_1'. The results of solving this problem are shown in Table I. Shown there are the optimum values of the parameters α, L, l', and ξ_3 for five fixed values of ξ_1' (variants $n=1,\dots,5$). It turned out that for $L>5$ cm, in all cases ξ_3' is essentially independent of L, so the data in the table are given for $L=5$ cm. It is clear from Table I that the best values of α and l' also change slowly as ξ_1' is raised. The optimum condition for the entire set of parameters corresponds to the tabulated data for $n=5$. These data and, in particular, $\xi_3=3.39\%$ are considerably better than the characteristics of the first maximum in the distribution of the product ozone concentration (1.905%) observed when hot oxygen is injected into the nozzle.

TABLE I. Optimizing the parameters of a nozzle ozonator.

n	ξ_1', %	α, cm^{-1}	l', cm	L, cm	ξ_3, %
1	0.50	0.63	0.64	5.0	0.49
2	1.00	0.64	0.64	5.0	0.96
3	2.00	0.65	0.71	5.0	1.85
4	4.00	0.67	0.73	5.0	3.18
5	4.92	0.70	0.95	5.0	3.39

Note that the main chemical processes responsible for the kinetics of ozone production under these conditions are the reactions

$$m_j=0.5 \quad m_j=2.0$$

$$\mathrm{O}+\mathrm{O}+M \rightarrow \mathrm{O}_2+M\ (M=\mathrm{O}_2)+1.9\%-3.7\%, \tag{1}$$

$$\mathrm{O}+\mathrm{O}_2+M \rightarrow \mathrm{O}_3+M\ (M=\mathrm{O}_2)-6.2\%+2.1\%, \tag{2}$$

$$\mathrm{O}+\mathrm{O}_2+M \rightarrow O_3+M\ (M=\mathrm{O})-0.2\%+0.3\%. \tag{3}$$

Here we have indicated the deviations in ξ_3 as a reaction rate constant of the form $K_j=m_jK_j^0$ is varied to the right of the reactions. The first column of numbers corresponds to $m_j=0.5$ (i.e., to a reduction in the rate constant by a factor of two) and the second, to $m_j=2$. When the other constants are varied, the deviation in the ozone concentration is much smaller.

Figure 2 shows the distribution of the concentration of

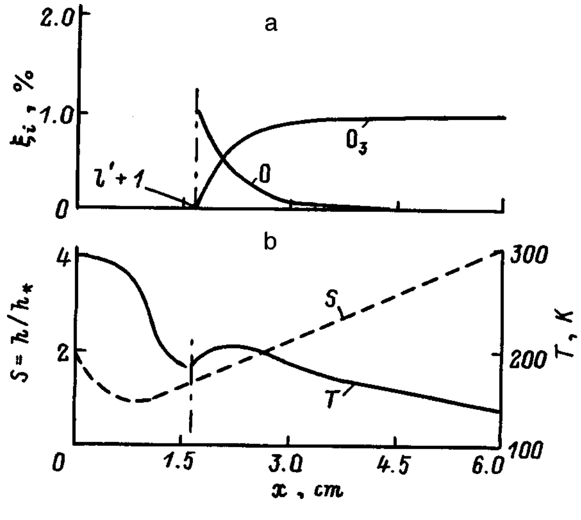

FIG. 2. Distributions of (a) the concentrations of the main components and (b) the gas temperature along the nozzle.

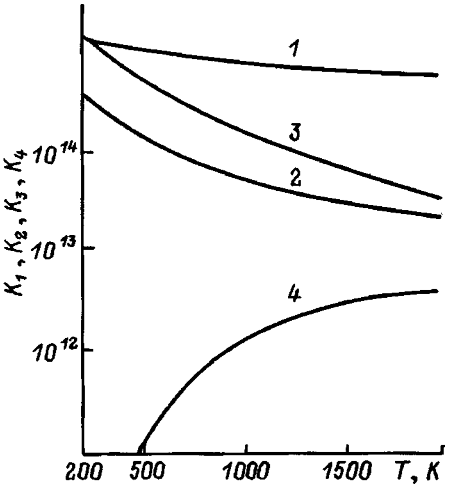

FIG. 3. The rate constants for reactions (*1*)–(*4*) as functions of gas temperature (K_1, K_2, and K_3 in cm^6/mol^{2}s, K_4 in cm^3/mol^{2}s).

the major components and temperature T along the nozzle S (dashed curve) for the case $n=2$ ($\xi_1'=1\%$) from Table I. The vertical line in Fig. 2 corresponds to the plane of the jump in the atomic O concentration. Under these conditions essentially all the excess concentration of O atoms goes into the formation of ozone molecules through recombination $O+O_2+M\rightarrow O_3+M$ and only a small fraction goes into the formation of O_2 molecules. It is clear that recombination, on the whole, leads to a small increase of 30–40 K in the temperature behind the jump, so that the ozone product molecules are not broken up.

The reasons for this behavior of the system are hidden in the way the rate constants for these reactions depend on the temperature (Fig. 3). It is clear from Fig. 3 (curve *4*) that the rate constant for loss of O_3 molecules in the exchange reaction

$$O+O_3\rightarrow O_2+O_2 \tag{4}$$

is small, while the rate constants for reactions (1) and (2), which remove O atoms as O_2 and O_3 molecules, are such that at low concentrations of atomic O, the main loss channel is reaction (2). Thus, for $T\approx 200$ K (the characteristic temperature of the gas in the nozzle for $T_0=300$ K), $K_1=1.72\times 10^{15}$ cm^6/mol^{2}s and $K_2=4.1\times 10^{14}$ cm^6/mol^{2}s. The rates of the corresponding reactions are $W_1\approx K_1\xi_1\xi_1$ and $W_2\approx K_2\xi_1\xi_2$, so

$$W_2/W_1=K_2\xi_1\xi_2/K_1\xi_1\xi_1=K_2/K_1\xi_1\cong 0.24/\xi_1 \tag{5}$$

(for $T=200$ K and $\xi_2=1$).

Therefore, for low atomic O concentrations, for example, when $\xi_1<0.024$, the ratio of the reaction rates $W_2/W_1\geqslant 10$, i.e., under these conditions almost all the O atoms are expended in ozone production.

PRODUCTION OF ATOMIC OXYGEN IN A NOZZLE OZONATOR WITH AN ELECTRICAL DISCHARGE

As shown in Ref. 4, the optimum from the standpoint of a maximum ozone yield in a nozzle ozonator with an electrical discharge is to generate the atomic oxygen in the supersonic section of the nozzle. Then, as the above analysis [see Eq. (5)] shows, in order for almost all the oxygen atoms formed in the discharge to be consumed in ozone production (the optimum regime), their concentration should not exceed $\xi_1\approx 0.024$. On the other hand, a large (by more than an order of magnitude) reduction in their concentration degrades the nozzle ozonator's productivity. We have, therefore, evaluated the feasibility of attaining these concentrations of O atoms in a discharge in the supersonic section of the nozzle.

The concentration (molar fraction) of atomic oxygen formed in a discharge in the region between the electrodes was estimated in the following way. In a gaseous discharge oxygen molecules are dissociated primarily by electron impact in the following reactions: direct electron impact dissociation of the molecule:

$$e+O_2\rightarrow e+O+O, \tag{6}$$

dissociation through excitation of electronic levels of the molecules lying above the dissociation threshold, with a subsequent transition to dissociating terms:

$$e+O_2\rightarrow e+O_2^*\rightarrow e+O+O, \tag{7}$$

where O_2^* is a molecule excited into a suitable electronic state, and dissociative attachment of electrons to oxygen molecules:

$$e+O_2\rightarrow O+O^-. \tag{8}$$

In order to find the rate constants for these reactions in the reduced-fiel d approximation, the electron energy distribution function $f(\varepsilon)$ was calculated from the Boltzmann equation[7]

$$\frac{E^2}{3}\frac{\varepsilon}{\sum_1 N_1\sigma_{ml}(\varepsilon)}\frac{df(\varepsilon)}{dt}+\sum_1\frac{2m}{M_1}N_1\varepsilon^2\sigma_{ml}(\varepsilon)\cdot\left[f(\varepsilon)+\frac{k_BT}{e}\frac{df(\varepsilon)}{d\varepsilon}\right]+\sum_1 N_1B\varepsilon\sigma_{\text{rot}}(\varepsilon)\cdot\left[f(\varepsilon)+\frac{k_BT}{e}\frac{df(\varepsilon)}{d\varepsilon}\right]=-\sum_1 N_1\sum_{ij}\int_{\varepsilon}^{\varepsilon+\varepsilon_{ij}}\sigma_{lij}(\varepsilon')\varepsilon' f(\varepsilon')d\varepsilon'. \tag{9}$$

Here E is the electric field strength, ε is the electron energy, σ_{ml} is the transport scattering cross section for electrons on oxygen atoms and molecules, N_1 is the concentration of molecules (atoms), m_1 and M_1 are the masses of an electron and of a molecule (atom), respectively, k_B is Boltzmann's constant, e is the electronic charge, B is the rotational constant, σ_{rot} are the cross sections for rotational excitation of the molecule, σ_{lij} are the cross sections for vibrational excitation of the molecules and electronic excitation of the atoms and molecules by electron impact, and ε_{ij} is the threshold for the corresponding excitation reaction. The cross sections were taken from the literature.[8–11] The correctness of the calculated $f(\varepsilon)$ was verified by the usual method[12] of comparing the variation in the drift velocity $v_{\text{dr}}(E/N)$ and characteristic electron temperature $T_{\text{ch}}(E/N)$ within the range of E/N considered here with published data.[13]

The rate constants for the reactions were found using the formula

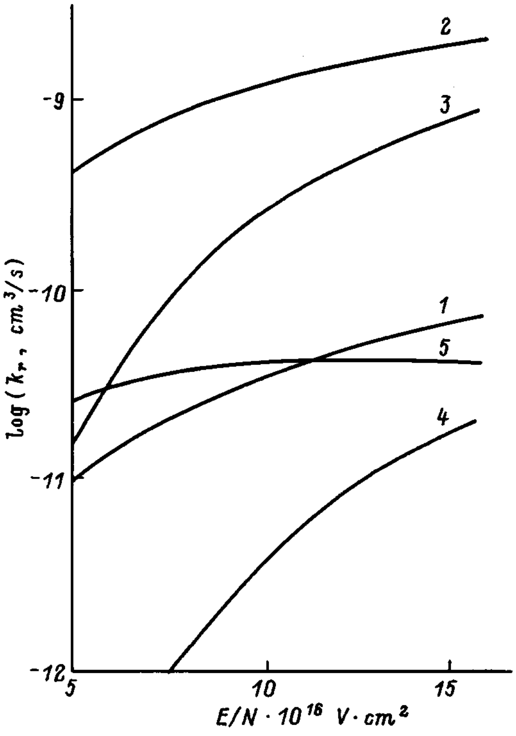

FIG. 4. Rate constants of the reactions leading to formation of atomic oxygen by electron impact as functions of E/N: curve *1*, reaction (6); curve *2*, reaction (7), $A^3\Sigma_u^+$ state; curve *3*, reaction (7), $B^1\Sigma_g^-$ state; curve *4*, reaction (7), the group of levels with a threshold of 9.9 eV; curve *5*, reaction (8).

$$k_{ij}(E/N)=\sqrt{2/m}\int_0^\infty \sigma_{ij}(\varepsilon)\varepsilon f(\varepsilon)d\varepsilon. \tag{10}$$

Since the electric field in the region between the electrodes in the supersonic section of the nozzle is nonuniform, calculations were done over a wide range of E/N, from 5×10^{-16} to 1.5×10^{-15} Vcm^2. The rate constants of reactions (6)–(8) for these measurement conditions are plotted in Fig. 4. The excited states corresponding to the different dissociation channels in reaction (7) are also indicated there.

The electron density n_e in the discharge region was estimated for typical values of the discharge current $I=25$–150 mA using Ohm's law

$$j=en_e v_{\text{dr}}, \tag{11}$$

where j is the current density.

For an electrode area of the order of, say, 2 cm^2, the range of variation in the electron density for the given range of E/N is roughly $n_e=3.4\times10^9$–4.0×10^{10} cm^{-3} according to these estimates.

The products of the reactions within the discharge are removed by the gas flow in the nozzle over short times $\tau=d/V\leqslant20$ μs (d is the size of the electrode along the flow and $V\sim500$ m/s is the flow velocity in the gap between the electrodes), so that the reverse reactions and the effect of vibrationally excited molecules were neglected in the calculations. Then the molecular fraction of atomic oxygen generated in the discharge can be estimated from the following expression:

$$\xi_1=2\int_0^t\sum_r k_r n_e \xi_2 dt+\int_0^t k_8 n_e \xi_2 dt, \tag{12}$$

where the sum over *r* refers to reactions (6) and (7), in which two oxygen atoms are formed, and k_8 is the rate constant for reaction (8).

The negative ion O$^-$ formed in reaction (8) does not directly lead to formation of an oxygen atom. Estimates using Eq. (12) give an atom concentration $[\text{O}]=5.3\times10^{14}$ -1.2×10^{16} cm^{-3} ($\xi_1=0.012$–0.262%). It should be noted that for specific calculations the Boltzmann equation (9) must be solved simultaneously with the balance equations for charged and excited particles, as has been done previously for nitrogen[14] and oxygen.[15] Nevertheless, these estimates showed that the molar fractions of atomic oxygen which can be obtained in a discharge in the supersonic section of the nozzle are close to the optimum values obtained above for a nozzle ozonator.

The presence of a second maximum in the ozone production curve corresponding to injecting a gas at temperature $T_0=300$ K into the nozzle, together with the above estimates of the concentration of O atoms in a discharge, has led to the initiation of a series of experiments in which molecular oxygen at room temperature is fed in from a nozzle in a gas dynamic generator, while the atoms required to produce ozone are created with the aid of an electrical discharge in the supersonic section of the flow, as proposed before in Refs. 4 and 5.

EXPERIMENT. TECHNIQUE AND RESULTS

The experiments were done on the apparatus illustrated schematically in Fig. 5. Molecular oxygen at a pressure of 1–10 atm was supplied from a cylinder to the inlet of a plane, wedge-shaped supersonic nozzle with a wedge angle of 10° and a throat height of 0.4 mm. The length and width of the nozzle were 100 and 15 mm, respectively. The outlet section of the nozzle was pumped constantly by a roughing pump with a speed of 20 liter/s.

A steady-state, transverse dc discharge was ignited a distance of 10 mm from the throat in the supersonic section of the nozzle. The upper and side walls of the nozzle were made of quartz glass and the lower wall, of steel. In the upper plane, 9 tungsten electrodes were mounted in staggered order, with each supplied by a separate ballast resistance connected to the power supply.

The discharge current was varied over $I=24$–150 mA, and the voltage to the electrodes was 0.8–1 kV. The molecular oxygen concentration in the supersonic section of the nozzle in the region between the electrodes ranged from 5.6×10^{18} to 4.4×10^{18} cm^{-3} along the gas flow.

The amount of ozone in the flow from the nozzle was determined by iodometric titration. Table II lists the measured ozone concentration for different discharge currents. One can see a linear dependence of the ozone yield on the energy input for moderate currents (70–80 mA), but a the discharge current is raised further, the ozone concentration saturates. The maximum ozone content in the flow under these conditions was 0.3%. The above estimates of the molar

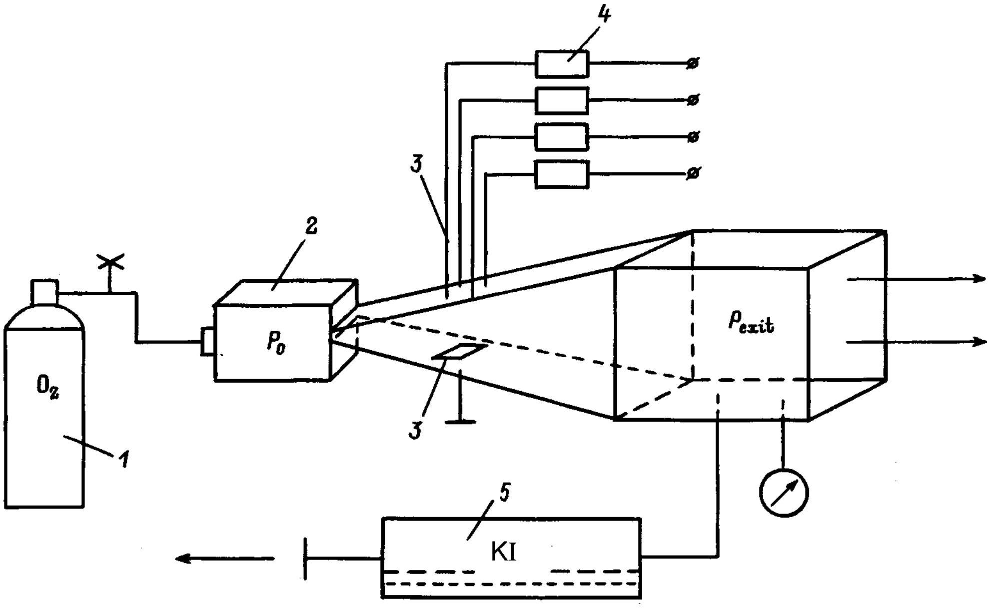

FIG. 5. Conceptual diagram of a nozzle ozonator employing an electrical discharge: *1* — cylinder of oxygen, *2* — nozzle ozonator, *3* — electrodes, *4* — ballast resistors, *5* — potassium iodide cell.

fraction of atomic oxygen ξ_1 formed in the discharge are also shown in Table II. Evidently, for high discharge currents condition (5) is satisfied, when essentially all of the oxygen atoms are expended in ozone formation. In this case, the measured and calculated molar fractions of ozone are in good agreement.

When the inlet pressure was changed from 1 to 10 atm, the relative concentration of ozone hardly changed. The oxygen feed rate for this nozzle is of order $1.57p_0$(atm) g/s. For an inlet gas pressure of $p_0 = 10$ atm, the feed rate is 15.7 g/s, so the ozone concentration is 6.4 g/m^3, and the measured and calculated molar fractions of ozone are in good agreement with each other.

It is clear that the productivity of the ozonator can be increased substantially by raising the initial gas pressure p_0. It should be noted that the reduced electric field between the electrodes corresponds roughly to the value where a large fraction of the energy of the discharge electrons is expended in reactions (7).[8]

CONCLUSION

In order to estimate the efficiency of the device described here, we have made a comparative analysis of industrial and semi-industrial equipment for ozone production. The generally accepted parameters characterizing the efficiency of ozonators are: the specific output per unit discharge area (g/h·dm^2), the specific energy expenditure for electrical synthesis (kW/h·kg), and the ozone yield per unit mass of the ozonator (g/kg). However, given the fundamental design differences between conventional ozonators, which customarily use a cylindrical discharge geometry, and the device described here, it is reasonable to introduce the specific ozone output per unit volume of the reactor, relative to the gas feed rate. Table III lists the comparative characteristics of several industrial and semi-industrial devices for ozone production. The best parameters at present are those of an rf ozonator under development at Moscow State University. It is evident from Table III that that ozonator is inferior to the one described here. The studies done in this paper show that it is possible to obtain a specific output per unit reactor volume exceeding that of the ozonator of Ref. 1 by a factor of 5. The specific output relative to the oxygen feed rate exceeds that of the ozonator of Ref. 1 by a factor of 20. The very high ozone yield per unit mass of the reactor is noteworthy.

TABLE II. Volume concentration of product ozone as a function of discharge power.

Run No.	I, mA	P, W	ξ_3, % (experiment)	ξ_1, % (calculation)
1	24	19	0.12	0.042
2	50	40	0.17	0.090
3	75	60	0.24	0.134
4	150	120	0.3	0.262

In the nozzle ozonator described here the calculated temperature in the reaction zone is -70 °C. It should be noted that in a nozzle, low temperatures are automatically maintained because of the flow is expanding. This makes it possible to avoid forced cooling, so the design of the ozonator is simpler and it is smaller. Furthermore, further expansion of the flow along the nozzle reduces the collision frequency of the molecules and, thereby, reduces the probability of dissociation of the ozone formed in the discharge. It should be noted, however, that there are at least two problems related to delivering the product ozone to the site where it will be used. First, there are reactions in the discharge afterglow, which can reduce its concentration. Second, as the flow is slowed down, the gas is heated and this leads to a drop in the ozone concentration.

Further improvement of the ozonator described here will involve optimizing the gas composition and the nozzle pro-

TABLE III. Comparative characteristics of some types of ozonators.

Type of resonator	Ozonator mass, kg	Volume, m^3	Gas feed rate, m^3/h	Ozone concentration, g/m^3	Ozone yield, g/h	Energy cost, kWh/kg	Specific yield per unit vol, $g^3/h \cdot m^3$	Specific yield relative to gas flow rate, $g/h \cdot m^6$	Ozone yield per unit mass of ozonator, g/kg	Frequency of discharge current, Hz
OPCh-61, USSR	1876	—	100	12.5	10 000	13	—	—	5.4	2400
OP 121, USSR	170	—	—	20	1600	22	—	—	0.95	50
USA Patent No. 4016060	—	—	—	14	450	2.2	—	—	—	10 000
‘‘Treilingas’’ Unit-76, France	3000	—	—	18	2016	12	—	—	0.67	1000
Single-tube rf ozonator, Moscow State Univ., Russia	25	1.4×10^{-3}	16.3	30	490	5.3	3.4×10^{5}	2.1×10^{4}	19.6	3000
Nozzle ozonator, Inst. of Mechanics, Moscow State Univ., Russia	0.2	1.5×10^{-5}	40	6.4	25	5.3	1.7×10^{6}	4.2×10^{5}	1700	dc discharge

file, as well as the type and energy characteristics of the electrical discharge.

The experimental results and numerical calculations presented here demonstrate the prospects for using supersonic nozzles in ozone production.

We sincerely thank V. A. Voblikova, V. G. Samoĭlovich, and V. A. Shakhatov for useful discussions and friendly support in the course of this work.

This work was partially supported by Grant No. 95-01-00415 of the Russian Fund for Fundamental Research.

[1] Yu. V. Filippov, V. A. Voblikova, and V. I. Panteleev, *Electrical Synthesis of Ozone* [in Russian], Izd. MGU, Moscow (1987), 237 pp.

[2] V. K. Dushin, I. E. Zabelinskiĭ, and O. P. Shatalov, ‘‘A method for obtaining superequilibrium ozone concentrations’’ [in Russian], Inventor’s Certificate No. 1230126, USSR; publ. Byull. Izobret., No. 48, p. 288 (1986).

[3] A. É. Krivonosova, V. N. Makarov, and O. P. Shatalov, in *Physicochemical Kinetics in Gas Dynamics* [in Russian], Izd. MGU, Moscow (1986), p. 122.

[4] V. N. Makarov and O. P. Shatalov, Izv. Akad. Nauk, Mekh. Zhidk. Gaza , No. 6, 139 (1994).

[5] V. N. Makarov and O. P. Shatalov, ‘‘A method for obtaining ozone in a supersonic flow’’ [in Russian], Russian Federation Patent for Invention No. 2057708, effective 10 April 1996.

[6] D. J. Wilde, *Optimum Seeking Methods* (Prentice-Hall, Englewood Cliffs, 1964; Nauka, Moscow, 1967, 267 pp.).

[7] W. L. Nighan, Phys. Rev. A **2**, 1989 (1970).

[8] R. Sh. Islamov, I. V. Kochetov, and V. G. Pevgov, FIAN Preprint No. 169 [in Russian], Fiz. Inst. Akad. Nauk SSSR, Moscow (1977).

[9] V. L. Bychkov and O. A. Gordeev, Khim. Fiz. **11**, 1064 (1992).

[10] Y. Itikawa, A. Ichimura, K. Onda *et al.*, J. Phys. Chem. Ref. Data **18**, 23 (1989).

[11] S. Geltman, J. Quant. Spectrosc. Radiat. Transf. **13**, 601 (1973).

[12] O. A. Gordeev and D. V. Khmara, Teplofiz. Vys. Temp. **32**, 133 (1994).

[13] J. Dutton, J. Phys. Chem. Ref. Data **4**, 577 (1975).

[14] A. V. Bodronosov, K. A. Vereshchagin, O. A. Gordeev *et al.*, Teplofiz. Vys. Temp. **34**, 666 (1996).

[15] A. B. Bessarabov and V. V. Rybkin, Teplofiz. Vys. Temp. **34**, 181 (1996).

Translated by D. H. McNeill

Poiseuille problem for an ellipsoidal-statistical equation and nearly specular boundary conditions

A. V. Latyshev and A. A. Yushkanov

Moscow Pedagogical Institute, Moscow, Russia

(Submitted May 12, 1997)
Zh. Tekh. Fiz. **68**, 27–31 (November 1998)

An analytical solution of the Poiseuille problem is obtained over a wide range of Knudsen numbers for the case when the tangential momentum accommodation coefficients of the channel walls are much less than unity. An expression for the mass flux is derived that is valid for Knudsen numbers much smaller than the reciprocal of the accommodation coefficients. A new intermediate flow regime, for which the mass flux is given by an expression that differs from the classical (macroscopic) form, is found to exist. © *1998 American Institute of Physics.* [S1063-7842(98)00511-X]

For the case when the tangential momentum accommodation coefficients of the channel walls are much smaller than unity, we have obtained an analytical solution over a wide range of Knudsen numbers. We have derived an expression for the mass flux in the channel as a function of the pressure gradient and the channel dimensions. This expression is valid for Knudsen numbers much smaller than the reciprocal of the accommodation coefficients. Thus, the standard macroscopic theory of gas flows in channels, valid for $\mathrm{Kn}\ll 1$, has been generalized to the case $1\ll \mathrm{Kn}\ll q$, where $q=\max(q_1,q_2)$. We show that there exists a new intermediate flow regime for $q\ll \mathrm{Kn}\ll 1$ in which the expression for the mass flux differs from the classical (macroscopic) form.

1. In recent years there has been heightened interest in problems regarding the motion of gases in channels.[1–4] As a rule, purely diffuse boundary conditions are considered. One exception is the case of the analytical solution to the problem of electron behavior in a metal layer for specular boundary conditions.[3,4] Purely specular boundary conditions, however, usually lead to a trivial result in the kinetic theory of gases. At the same time, interest in problems with a restricted geometry is currently increasing. Thus, the plane Poiseuille problem has been examined[5] in an almost continuous regime with first- and second-order slip. The Poiseuille problem for a cylindrical pipe has been solved numerically.[6] Gas flows in a layer have been examined[7], as in Ref. 6, by expanding in a Neumann series. In all these papers, purely diffuse boundary conditions were used. The opposite case of nearly specular boundary conditions has not been studied. The purpose of the present paper is to fill this gap.

Two methods, the Wiener–Hopf and Case methods,[8] are customarily used for solving boundary value problems analytically. The latter has the advantage of providing a distribution function, as well as the values of the macroscopic parameters. In this paper we use Case's method. In Refs. 3 and 4, the Case approach was used to develop a method for solving kinetic problems for layers with specular boundary conditions. In this paper, this method is generalized to the case of nearly specular boundary conditions. We consider the classical Poiseuille problem of a gas flow in a plane channel driven by a pressure gradient. An analytical solution of the problem is obtained over a wide range of Knudsen numbers for tangential momentum accommodation coefficients much smaller than unity. An ellipsoidal-statistical model for the kinetic equation is used which yields the correct Prandtl number.

2. Let us consider a plane channel of width $L=2d$, in which a longitudinal pressure gradient is maintained. We shall assume that the gas flow in the channel is stationary. We shall describe the kinetics of the process using a Boltzmann equation with an ellipsoidal model for the collision integral.[9] Let us introduce a system of coordinates centered on the middle of the channel with the x axis perpendicular to the wall. Let the z axis be directed along the pressure gradient. We assume that the process is isothermal. We shall assume that the relative pressure drop over the mean free path is small. In this case, the problem can be linearized and a distribution function of the form $f=f_0(1+h)$ can be sought. Here

$$f_0=n(z)(m/2kT)^{3/2}\exp[-mv^2/2kT],$$

where n is the concentration of gas molecules, m is the molecular mass, T is the temperature, $\mathbf{v}$ is the molecular velocity, k is the Boltzmann constant, and the function h is a linear correction to the local equilibrium function f_0.

For this problem, the ellipsoidal-statistical equation can be written in the form

$$v_x\frac{\partial h}{\partial x}+v_z\frac{\partial \ln n}{\partial z}=\nu\left[\frac{m}{kT}v_zU-2\left(\frac{m}{2kT}\right)^2 v_xv_zP_{xz}-h\right]. \qquad (1)$$

Here

$$U=\left(\frac{m}{2\pi kT}\right)^{3/2}\int\exp\left[-\frac{mv^2}{2kT}\right]v_zh\,d^3v,$$

$$P_{xz}=\left(\frac{m}{2\pi kT}\right)^{3/2}\int\exp\left[-\frac{mv^2}{2kT}\right]v_xv_zh\,d^3v,$$

and ν is a constant having the physical significance of a collision frequency.

We introduce the dimensionless variables $\mathbf{C}=\sqrt{\beta}\mathbf{v}$, $U'=\sqrt{\beta}U$, $x'=\sqrt{\beta}\nu x$, $d'=\sqrt{\beta}d\nu$, and $\beta=m/2kT$. In the following we shall omit the primes on U', x', and d'. Then Eq. (1) can be rewritten in the form

$$C_x\frac{\partial h}{\partial x}+h+C_zK=2\pi^{-3/2}\int e^{-C'^2}C_zC_z'(1-C_xC_x')hd^3C', \tag{2}$$

where

$$K=\frac{\partial \ln n}{\partial z}.$$

The diffuse–specular boundary conditions on the function h have the following form:

$$h(-d,\mathbf{C})=(1-q_1)h(-d,\mathbf{C}-2\mathbf{n}_1,\mathbf{C}),\quad C_x>0,$$

$$h(d,\mathbf{C})=(1-q_2)h(d,\mathbf{C}-2\mathbf{n}_2\mathbf{C}),\quad C_x<0. \tag{3}$$

Here q_1 and q_2 are the tangential momentum accommodation coefficients (the specular coefficients) for the lower and upper surfaces, respectively, and $\mathbf{n}_1$ and $\mathbf{n}_2$ are the unit vectors normal to the lower and upper surfaces, directed toward the interior of the channel. It is clear from the form of Eq. (2) and the boundary conditions (3) that h can be sought in the form $h=v_z\psi(x,\mu)$, where $\mu=v_x$. Then Eq. (3) can be transformed to the equation

$$\mu\frac{\partial\psi}{\partial x}+\psi(x,\mu)+K=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}e^{-\mu'^2}(1-\mu\mu')\psi(x,\mu')d\mu'. \tag{4}$$

The boundary conditions (3) for the function ψ can be rewritten as

$$\psi(-d,\mu)=(1-q_1)\psi(-d,-\mu),\quad \mu>0,$$

$$\psi(d,\mu)=(1-q_2)\psi(d,-\mu),\quad \mu<0. \tag{5}$$

It is easy to verify that the particular solution of Eq. (4) is the function

$$\psi_0(x,\mu)=K\left[\frac{3}{2}x^2-2x\mu+2\mu^2+a_0+a_1\left(x-\frac{2}{3}\mu\right)\right],$$

where a_0 and a_1 are arbitrary constants.

3. We shall seek a solution of the homogeneous equation corresponding to Eq. (4) in the form

$$\psi_\eta(x,\mu)=\exp\left(-\frac{x}{\eta}\right)\Phi(\eta,\mu)$$

with the condition

$$\int_{-\infty}^{\infty}e^{-\mu'^2}\Phi(\eta,\mu')d\mu'\equiv 1. \tag{6}$$

We obtain the characteristic equation

$$(\eta-\mu)\Phi(\eta,\mu)=\frac{1}{\sqrt{\pi}}\eta. \tag{7}$$

For $-\infty<\eta<+\infty$, we take a solution of Eqs. (6) and (7) in the space of generalized functions[10]

$$\Phi(\eta,\mu)=\frac{1}{\sqrt{\pi}}\eta P\frac{1}{\eta-\mu}+e^{\eta^2}\lambda(\eta)\delta(\eta-\mu).$$

Here the symbol Px^{-1} denotes the distribution — the principal value of the integral of x^{-1}, $\delta(x)$ is the Dirac delta function, and $\lambda(z)$ is the dispersion function given by

$$\lambda(z)=1+\frac{1}{\sqrt{\pi}}z\int_{-\infty}^{\infty}e^{-\tau^2}\frac{d\tau}{\tau-z}.$$

We construct the general solution of the homogeneous equation corresponding to Eq. (4), in the form of an expansion in the eigenfunctions $\Phi(\eta,\mu)$, of the characteristic equation (7),

$$\psi_c(x,\mu)=\int_{-\infty}^{\infty}e^{-x/\eta}\Phi(\eta,\mu)a(\eta)d\eta, \tag{8}$$

where the function $a(\eta)$ is referred to as the continuum spectrum coefficient.

Let us substitute the general solution of Eq. (4) $\Psi(x,\mu)=\psi_0(x,\mu)+\psi_c(x,\mu)$ in the boundary conditions (5). We then obtain

$$\begin{aligned}[\psi_c(-d,\mu)-\psi_c(-d,-\mu)]&+q_1\psi_c(-d,-\mu)\\ &=-[\psi_0(-d,\mu)-\psi_0(-d,-\mu)]\\ &\quad -q_1\psi_0(-d,-\mu),\quad \mu>0,\end{aligned} \tag{9}$$

$$\begin{aligned}[\psi_c(d,\mu)-\psi_c(d,-\mu)]&+q_2\psi_c(d,-\mu)\\ &=-[\psi_0(d,\mu)-\psi_0(d,-\mu)]-q_2\psi_0(d,-\mu),\\ \mu<0.&\end{aligned} \tag{10}$$

Note that

$$\psi_0(d,\mu)-\psi_0(d,-\mu)=2K\left(-2d\mu-\frac{2}{3}a_1\mu\right),$$

$$\begin{aligned}\psi_c(x,\mu)-\psi_c(x,-\mu)=\int_{-\infty}^{\infty}&\Phi(\eta,\mu)[e^{-x/\eta}a(\eta)\\ &-e^{x/\eta}a(-\eta)]d\eta.\end{aligned}$$

From the conditions (9) and (10), we obtain the integral equations

$$\begin{aligned}\int_{-\infty}^{\infty}&\Phi(\eta,\mu)[e^{d/\eta}a(\eta)-e^{-d/\eta}a(-\eta)]d\eta\\ &+q_1\int_{-\infty}^{\infty}e^{-d/\eta}\Phi(\eta,\mu)a(-\eta)d\eta\\ &=\mu\varphi_1(d,q_1)-q_1\varphi_2(\mu,d),\quad \mu>0,\end{aligned} \tag{11}$$

$$\int_{-\infty}^{\infty}\Phi(\eta,\mu)[e^{-d/\eta}a(\eta)-e^{d/\eta}a(-\eta)]d\eta$$

$$+q_2\int_{-\infty}^{\infty}e^{d/\eta}\Phi(\eta,\mu)a(-\eta)d\eta$$

$$=\mu\varphi_1(-d,q_2)-q_2\varphi_2(\mu,-d),\quad \mu<0. \tag{12}$$

Here

$$\varphi_1(d,q_i)=2K\left(2d+\frac{2}{3}a_1+q_id-\frac{1}{3}q_ia_1\right),$$

$$\varphi_2(\mu,d)=K\left(\frac{3}{2}d^2+a_0-a_1d+2\mu^2\right).$$

The first integrals on the left-hand sides of Eqs. (11) and (12) are odd functions of μ. Let us assume that q_1 and q_2 are small, i.e., $q_i\ll 1$, where $i=1,2$. For a wide channel, where $\mathrm{Kn}=l/2d\ll 1$ (l is the mean free path of the molecules), the second integrals on the left of Eqs. (11) and (12) can be neglected. In fact, the two integrals are related through the collision integral of Eq. (4), from which it is clear that their ratio is of order Kn. Thus, the second integrals in Eqs. (11) and (12) are of order $q_i\mathrm{Kn}$ relative to the first integrals. We shall assume that the condition $q_i\mathrm{Kn}\ll 1$, with $i=1,2$, is satisfied. In this case, the second integrals on the left of Eqs. (11) and (12) can, as we have said, be neglected. Thus, in the following we shall consider the equations

$$\int_{-\infty}^{\infty}\Phi(\eta,\mu)[e^{d/\eta}a(\eta)-e^{-d/\eta}a(-\eta)]d\eta$$

$$=\mu\varphi_1(d,q_1)-q_1\varphi_2(\mu,d),\quad \mu>0,$$

$$\int_{-\infty}^{\infty}\Phi(\eta,\mu)[e^{-d/\eta}a(\eta)-e^{d/\eta}a(-\eta)]d\eta$$

$$=\mu\varphi_1(-d,q_2)-q_2\varphi_2(\mu,-d),\quad \mu<0.$$

The left-hand sides of these equations are odd functions. Let us extend both equations over the entire axis, continuing their right-hand sides as odd functions. We obtain the equations

$$\int_{-\infty}^{\infty}\Phi(\eta,\mu)[e^{d/\eta}a(\eta)-e^{-d/\eta}a(-\eta)]d\eta$$

$$=\mu\varphi_1(d,q_1)-q_1\varphi_2(\mu,d)\,\mathrm{sign}\,\mu,\quad -\infty<\mu<+\infty,$$

$$\int_{-\infty}^{\infty}\Phi(\eta,\mu)[e^{-d/\eta}a(\eta)-e^{d/\eta}a(-\eta)]d\eta=\mu\varphi_1(-d,q_2)$$

$$+q_2\varphi_2(\mu,-d)\,\mathrm{sign}\,\mu,\quad -\infty<\mu<+\infty.$$

4. Substituting the eigenfunctions $\Phi(\eta,\mu)$ in these equations and introducing the two auxiliary functions

$$N(z)=\int_{-\infty}^{\infty}\eta[e^{d/\eta}a(\eta)-e^{-d/\eta}a(-\eta)]\frac{d\eta}{\eta-z}, \tag{13}$$

$$M(z)=\int_{-\infty}^{\infty}\eta[e^{-d/\eta}a(\eta)-e^{d/\eta}a(-\eta)]\frac{d\eta}{\eta-z}, \tag{14}$$

we arrive at two Riemann boundary value problems

$$\lambda^+(\mu)[N^+(\mu)-\sqrt{\pi}\varphi_1(d,q_1)\mu]-\lambda^-(\mu)[N^-(\mu)$$

$$-\sqrt{\pi}\varphi_1(d,q_1)\mu]=-q_1\varphi_2(\mu,d)\,\mathrm{sign}\,\mu,$$

$$-\infty<\mu<+\infty, \tag{15}$$

$$\lambda^+(\mu)[M^+(\mu)-\sqrt{\pi}\varphi_1(-d,q_2)\mu]-\lambda^-(\mu)[M^-(\mu)$$

$$-\sqrt{\pi}\varphi_1(-d,q_2)\mu]=q_2\varphi_2(\mu,-d)\,\mathrm{sign}\,\mu,$$

$$-\infty<\mu<+\infty. \tag{16}$$

Given the behavior of the functions in Eqs. (15) and (16), we obtain the general solutions of these problems,

$$N(z)=2\sqrt{\pi}K\left[-2d+\frac{2}{3}a_1+q_1\left(d-\frac{1}{3}a_1\right)\right]z$$

$$-q_1K\left(\frac{3}{2}d^2+a_0-a_1d\right)\frac{\psi(z)}{\lambda(z)}-2q_1K\frac{z+z^2\psi(z)}{\lambda(z)}, \tag{17}$$

$$M(z)=2\sqrt{\pi}K\left[2d+\frac{2}{3}a_1-q_2\left(d+\frac{1}{3}a_1\right)\right]z$$

$$+q_2K\left(\frac{3}{2}d^2+a_0-a_1d\right)\frac{\psi(z)}{\lambda(z)}+2q_2K\frac{z+z^2\psi(z)}{\lambda(z)}. \tag{18}$$

Here

$$\psi(z)=\int_{-\infty}^{\infty}(\mathrm{sign}\,\mu)\mu e^{-\mu^2}\frac{d\mu}{\mu-z}.$$

Note that N, M, and ψ are odd functions. In order for the solutions (17) and (18) to serve as the auxiliary functions N and M introduced in Eqs. (13) and (14), we eliminate the simple pole at the point $z=\infty$ in the solutions (17) and (18). This is done by choosing a_0 and a_1 from the equations

$$\sqrt{\pi}\left[-d(2-q_1)+\frac{a_1}{3}(2-q_1)\right]$$

$$-q_1\left(\frac{3}{2}d^2+a_0-a_1d+2\right)=0,$$

$$\sqrt{\pi}\left[d(2-q_2)+\frac{a_1}{3}(2-q_2)\right]$$

$$+q_2\left(\frac{3}{2}d^2+a_0+a_1d+2\right)=0.$$

From these equations we find

$$a_0=-\frac{4\sqrt{\pi}d+(q_1+q_2)\left(2+\dfrac{9}{2}d^2-2\sqrt{\pi}d\right)+q_1q_2\left(\dfrac{3}{\sqrt{\pi}}d-1\right)\left(2+\dfrac{3}{2}d^2-\sqrt{\pi}d\right)}{\left(\dfrac{3}{\sqrt{\pi}}d-1\right)q_1q_2+q_1+q_2},$$

$$a_1=-\frac{3d(q_1-q_2)}{\left(\dfrac{3}{\sqrt{\pi}}d-1\right)q_1q_2+q_1+q_2}.$$

5. We now calculate the mass flux in the channel,

$$J=2d\int_{-d}^{d}j(x)dx,$$

where

$$j(x)=\int m\varphi f_0 v_y^2 d^3v$$

$$=\rho\frac{1}{\sqrt{\pi\beta}}\frac{1}{2}\int_{-\infty}^{\infty}e^{-\mu^2}\psi(x,\mu)d\mu=\frac{\rho}{\sqrt{\beta}}J_1$$

is the mass flux density and

$$J_1=\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty}e^{-\mu^2}\psi(x,\mu)d\mu$$

is the dimensionless mass flux density of the gas.

We have

$$\int_{-\infty}^{\infty}e^{-\mu^2}\psi(x,\mu)d\mu=\int_{-\infty}^{\infty}e^{-\mu^2}\psi_0(x,\mu)d\mu$$

$$+\int_{-\infty}^{\infty}e^{-\mu^2}d\mu\int_{-\infty}^{\infty}e^{-x/\eta}\Phi(\eta,\mu)d\eta$$

$$=\sqrt{\pi}K\left(\frac{3}{2}x^2+1+a_0+a_1x\right)$$

$$+\int_{-\infty}^{\infty}e^{-x/\eta}a(\eta)d\eta.$$

Therefore, the mass flux of the gas is

$$J=\frac{2d\rho}{\sqrt{\pi}}K\left[\frac{1}{2}d^3+(1+a_0)d\right]+\frac{2d\rho}{\sqrt{\pi\beta}}\int_{-\infty}^{\infty}\eta\sinh\frac{d}{\eta}a(\eta)d(\eta).$$

From Eqs. (13), (17), and (14), (18), respectively, we have

$$2\pi i\eta(e^{d/\eta}a(\eta)-e^{-d/\eta}a(-\eta))$$

$$=-q_1K\left(\frac{3}{2}d^2+a_0-a_1d+2\eta^2\right)\left(\frac{\psi^+(\eta)}{\lambda^+(\eta)}\right.$$

$$\left.-\frac{\psi^-(\eta)}{\lambda^-(\eta)}\right)-2q_1K\eta\left(\frac{1}{\lambda^+(\eta)}-\frac{1}{\lambda^-(\eta)}\right),$$

$$2\pi i\eta(e^{-d/\eta}a(\eta)-e^{d/\eta}a(-\eta))$$

$$=-q_2K\left(\frac{3}{2}d^2+a_0+a_1d+2\eta^2\right)\left(\frac{\psi^+(\eta)}{\lambda^+(\eta)}-\frac{\psi^-(\eta)}{\lambda^-(\eta)}\right)$$

$$-2q_2K\eta\left(\frac{1}{\lambda^+(\eta)}-\frac{1}{\lambda^-(\eta)}\right).$$

Subtracting the second equation from the first, we find

$$4\pi i\eta[a(\eta)+a(-\eta)]\sinh\frac{d}{\eta}=-2K\eta\left(\frac{1}{\lambda^+(\eta)}-\frac{1}{\lambda^-(\eta)}\right)$$

$$\times(q_1+q_2)+K\left(\frac{\psi^+(\eta)}{\lambda^+(\eta)}-\frac{\psi^-(\eta)}{\lambda^-(\eta)}\right)$$

$$\times\left[a_1d(q_1-q_2)-\left(\frac{3}{2}d^2+a_0+2\eta^2\right)(q_1+q_2)\right]. \tag{19}$$

Note that

$$\frac{\psi^+(\eta)}{\lambda^+(\eta)}-\frac{\psi^-(\eta)}{\lambda^-(\eta)}=2\sqrt{\pi}i\frac{|\eta|e^{-\eta^2}t(-|\eta|)}{\lambda^+(\eta)\lambda^-(\eta)},$$

$$\frac{1}{\lambda^+(\eta)}-\frac{1}{\lambda^-(\eta)}=-\frac{2\sqrt{\pi}i\eta e^{-\eta^2}}{\lambda^+(\eta)\cdot\lambda^-(\eta)},$$

where

$$t(\eta)=\int_{-\infty}^{\infty}e^{-\tau^2}\frac{d\tau}{\tau-\eta}. \tag{20}$$

Here we use the notation

$$\gamma_k=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{|\eta|^k e^{-\eta^2}t(-|\eta|)d\eta}{|\lambda^+(\eta)|^2}\ (k=1,3),$$

$$\delta=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\eta^2e^{-\eta^2}d\eta}{|\lambda^+(\eta)|^2}. \tag{21}$$

Numerical estimates show that $\gamma_1\approx1/\sqrt{\pi}$ and $\gamma_3\approx3/(2\sqrt{\pi})$, and from the contour integration we find that $\delta\approx3/\sqrt{\pi}$. Using Eq (20) with the notation (21), after integrating Eq. (19) from $-\infty$ to $+\infty$, we obtain

$$\int_{-\infty}^{\infty}\eta a(\eta)\sinh\frac{d}{\eta}\,d\eta=\sqrt{\pi}K\left\{q\delta-\frac{1}{2}\left[\left(\frac{3}{2}d^2+a_0\right)\times\gamma_1+2\gamma_3\right]q+\frac{1}{2}a_1dp\gamma_1\right\},$$

where $q=(q_1+q_2)/2$ and $p=(q_1-q_2)/2$.

Therefore, the mass flux is given by

$$J=\frac{2d\rho}{\sqrt{\pi}}K\left[\frac{1}{2}d^3+(1+a_0)d+q(\delta-\gamma_3)-q\frac{1}{2}\left(\frac{3}{2}d^2+a_0\right)\gamma_1+\frac{1}{2}pa_1d\gamma_1\right]$$

or, in dimensional form,

$$J=3\frac{\eta}{P}\left[\frac{1}{2}\left(\frac{d}{l}\right)^3+(1+a_0)\frac{d}{l}+q(\delta-\gamma_3)-\frac{1}{2}q\left(\frac{3}{2}\left(\frac{d}{l}\right)^2+a_0\right)\gamma_1+\frac{1}{2}pa_1\frac{d}{l}\gamma_1\right]|\nabla P|, \qquad (22)$$

where P is the gas pressure.

Let us consider the case of a wide channel, with a Knudsen number $\mathrm{Kn}=l/2d\ll 1$. We shall assume that the accommodation coefficients q_1 and q_2 are of the same order of magnitude and, as before, write $q=(q_1+q_2)/2$. Two flow regimes are possible in a wide channel. The first corresponds to the case $\mathrm{Kn}\ll q$. Then the expression for the mass flux takes the form (in dimensional variables)

$$J=-\frac{2\rho}{3\eta}d^3L\frac{dP}{dz},$$

where L is the transverse dimension of the channel.

This formula is the same as the standard formula for the mass flux in a channel in the hydrodynamic limit.[11]

There is another gas flow regime, when $q\ll\mathrm{Kn}\ll 1$. In this case, the expression for the mass flux of gas takes the form

$$J=-\frac{8}{q_1+q_2}\frac{\rho}{\eta}Ld^2l\frac{dP}{dz}.$$

Therefore, for nearly specular boundary conditions, there is a regime in which, despite the smallness of the Knudsen number, the expression for the mass flux of gas in a channel differs from the hydrodynamic formula. The transition to the purely hydrodynamic regime occurs under the stronger condition $\mathrm{Kn}\ll q$.

This work was supported by the Russian Fund for Fundamental Research (Project No. 97-01-00333).

[1] C. Cercignani, *Theory and Application of the Boltzmann Equation* (American Elsevier, New York, 1975; Mir, Moscow, 1978, 495 pp.).

[2] J. L. Lebowitz and E. W. Montroll (Eds.), *Nonequilibrium Phenomena I: The Boltzmann Equation* (North-Holland, Amsterdam–New York, 1983; Mir, Moscow, 1986, 272 pp.).

[3] A. V. Latyshev, A. G. Lesskis, and A. A. Yushkanov, Teor. Mat. Fiz. **90**, 179 (1992).

[4] A. V. Latyshev and A. A. Yushkanov, Zh. Vychisl. Mat. Mat. Fiz. **33**, 259 (1993).

[5] S. K. Loyalka and K. A. Hickey, Physica A **160**, 395 (1989).

[6] S. K. Loyalka and S. K. Hamoody, Phys. Fluids **2**, 2061 (1990).

[7] M. Hasegawa and Y. Sone, Phys. Fluids **3**, 466 (1991).

[8] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967; Mir, Moscow, 1972, 384 pp.).

[9] L. H. Holway, Phys. Fluids **9**, 1658 (1966).

[10] V. S. Vladimirov, *Generalized Functions in Mathematical Physics*, Mir, Moscow (1979), 280 pp.

[11] L. G. Loĭtsyanskiĭ, *The Mechanics of Liquids and Gases* (Nauka, Moscow, 1973), 497 pp.

Translated by D. H. McNeill

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Nonstationary interaction of a supersonic flying object with extended low-density regions in the atmosphere

O. M. Velichko, V. D. Urlin, and B. P. Yakutov

Russian Federal Nuclear Center, 607190 Saratov, Nizhniĭ Novgorod Region, Russia

(Submitted June 26, 1997)
Zh. Tekh. Fiz. **68**, 32–37 (November 1998)

A three-dimensional numerical model is developed and the aerodynamic loads are determined for the three-dimensional flow acting on a blunt cone flying in the atmosphere along the boundary of a rarefied region, as functions of the degree of immersion of the surface of the body in low-density air. The nonstationary gas- and aerodynamic processes accompanying the entry and exit of the body from the rarefied region are studied. The aerodynamic coefficients obtained with an approximate model are compared with the three-dimensional calculations and found to be in satisfactory agreement with them. © *1998 American Institute of Physics.* [S1063-7842(98)00611-4]

INTRODUCTION

The motion of bodies through density inhomogeneities in the atmosphere has been studied in various gasdynamic formulations.[1–7] The interest in this problem is explained by the fact that when bodies interact with atmospheric inhomogeneities, aerodynamic loads develop which can have a significant effect on the motion of these bodies, by changing the drag coefficient or by creating an additional lift force and rotational moment. During supersonic motion of bodies in inhomogeneous media, substantial changes can also occur in the bow shock and in the structure of the shock layer, which are of independent interest. Most papers devoted to the interaction with inhomogeneities consider inhomogeneities in the form of rarefied regions created by the delivery of thermal energy to the gas. In most cases it is assumed that an interaction of the body with a rarefied region does not change the axial symmetry of the flow around the body.[1–3,5,6] References 4 and 7, however, began the study of this sort of interaction between a conical shock wave and a rarefied region in which the initial axial symmetry of the flow around a blunt cone is broken. In Ref. 7 it was assumed that the rarefied region which comes into contact with the shock has the shape of a channel whose axis is orthogonal to that of the cone. In this paper we examine another variant of this interaction, of which a particular example was first studied in Ref. 4. We study the case in which the rarefied volume is extended in the direction of motion of the body and the cone moves along the boundary separating the regions of different density. Here part of the surface of the body and the bow shock are immersed in low-density air and move in it. The aerodynamic loads experienced by the body during this sort of motion can be substantial. This is suggested in our earlier work,[4] in which we have examined the case where a half cone is immersed in rarefied air. In this regard, it is interesting to track how the character of the flow and aerodynamic forces change with different degrees of immersion of a cone in a rarefied volume. In this paper, numerical simulation of a three-dimensional flow is used to study the nonstationary interaction of a conical body with a rarefied volume for different orientations of the body and volume. Because the rarefied region is not positioned symmetrically with respect to the axis of the wave in this statement of the problem, the resulting flow has a pronounced three-dimensional, nonstationary character. The most natural method of studying this kind of flow is three-dimensional numerical modeling. Here we also present some calculations of the aerodynamic characteristics of a blunt cone moving along the boundary between regions with different densities.

STATEMENT OF THE PROBLEM

We consider a conical shock wave created by a blunt cone flying with velocity U_∞ in a homogeneous atmosphere at zero angle of attack. At time $t=0$, a portion of the surface of the bow shock comes into contact with a rarefied region of the atmosphere. The low-density region is isolated in space by two plane contact boundaries. The axis of the cone is parallel to one of the contact boundaries and perpendicular to the other. The characteristic size Δ of the rarefied region in the direction of motion of the body is several times greater than the length L of the cone. The pressure P_0 in the rarefied region equals the pressure P_∞ in the unperturbed atmosphere, while its density ρ_0 is lower than the unperturbed density ρ_∞. Figure 1 shows the locations of the conical shock, the cone, and the rarefied region relative to one another, as well as the coordinate system used in the calculations. The coordinate of the contact surface parallel to the axis of the cone and the YOZ plane equals X, so that the distance from the axis of the cone to the boundary of the rarefied region is $|X|$. We introduce the dimensionless impact parameter $\varepsilon=X/R_m$ which is convenient for determining the position of the body relative to the rarefied volume (R_M is the radius of the central cross section of the cone). When $\varepsilon=-1$ the cone is entirely in dense air, while when $\varepsilon=1$ it is entirely in rarefied air. If $\varepsilon=0$, then half the body will be in rarefied air and the other half in dense air. In the calculations ε was varied over the limits $-1\leqslant\varepsilon\leqslant 1$. The purpose of the

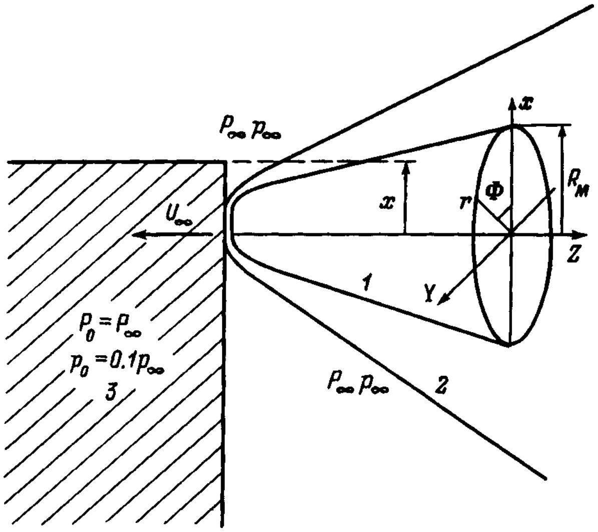

FIG. 1. A sketch of the interaction of a body and a conical shock wave with a rarefied region: (*1*) blunt cone, (*2*) bow shock wave, (*3*) low-pressure region.

calculations was to establish how the aerodynamic forces acting on the body depend on the degree of immersion of the cone in the rarefied volume or, in other words, to determine the dependence of the aerodynamic coefficients on the "impact parameter" ε, i.e., the distance between the axis of the body and the boundary of the rarefied region.

COMPUTATIONAL TECHNIQUE

The problem of the interaction of a rarefied volume with a conical bow shock was solved by a method described in the literature.[7,9,10] The gas flowing past the object is assumed to be inviscid and thermally nonconducting, and its motion is described by a three-dimensional, nonstationary system of Euler equations. The form of the equations used in the present calculations is given in Ref. 7.[1] The equation of state of an ideal gas with a constant adiabatic index of $\gamma=1.4$ was used in the calculations. The numerical solution of the problem is carried out in a cylindrical coordinate system rigidly attached to the body (Fig. 1). The Z axis coincides with the axis of the cone. The flow is symmetric with respect to the XOZ plane. The angle φ is reckoned from the upper half of the XOZ plane, viewing from the direction of the nose. The rarefied region is modeled by specifying time dependent boundary conditions at the surface of the bow shock. The coordinates of the segment of the outer surface of the shock which is in contact with the rarefied region at time t are given by

$$Z_0 \leqslant Z \leqslant Z_0 + U_\infty t, \quad r_W(Z,\varphi)\cos\varphi \leqslant X. \tag{1}$$

In this segment the parameters of the air are

$$U_0 = U_\infty, \quad P_0 = P_\infty, \quad \rho_0 < \rho_\infty. \tag{2}$$

In Eqs. (1) and (2) Z_0 is the coordinate of the boundary of the volume perpendicular to the axis of the cone at time $t=0$, X is the coordinate of the contact boundary parallel to the axis of the cone, $r_W(Z,\varphi)$ is the radius of the bow shock at the cross section with coordinate Z and direction specified by the angle φ, the subscript 0 denotes parameters of the air on the segment of the bow shock lying within the rarefied region, and the parameters of the air on the rest of the shock surface are equal to the unperturbed P_∞, ρ_∞, and U_∞.

The three-dimensional nonstationary Euler equations are integrated using a Godunov explicit difference scheme with first-order accuracy.[8] The calculations were done with explicit capturing of the bow shock. Along the Z axis the computational region contains 120 mesh points, which become denser near the blunted part. The computational region is divided uniformly in the angular coordinate. The total number of mesh points in the computational region for the angle φ over 0 to 180° is 31. Between the surfaces of the body and the bow shock, the coordinate is uniformly divided into 20 mesh points. Thus, the total number of mesh points in the computational region is $120\times31\times20=74\,400$. The finite difference method for calculating shock wave flows on meshes of this type using a Godunov scheme has been described elsewhere.[7,9,10]

COMPUTATIONAL RESULTS

Dimensionless variables are used to represent the computational results. The coordinates are in units of the length L of the cone, the pressure in units of P_∞, the density in units of ρ_∞, the velocity in units of U_∞, and the time in units of L/U_∞. In all versions of the three-dimensional calculation, the velocity U_∞ of the cone remained constant and corresponded to a Mach number $\mathrm{Ma}_\infty=15.6$. The inclination of the side surface of the cone to the axis was 7°. The parameter ε was varied and took the values -0.85, -0.54, 0, and $+0.54$. The density of the air in the rarefied volume was 10 times smaller than the unperturbed density $\rho_0=0.1$. The distance Z_0 in a specific calculation depended on the magnitude of ε and was determined from the coordinates of the unperturbed shape of the bow shock. The calculations were begun at time $t=0$, which corresponded to the beginning of contact between the rarefied region and the bow shock and were ended when the body had penetrated the rarefied region by a distance of 1–2 times its own length, i.e., when $t \sim (1-2)L/U_\infty$. In the calculations with $\varepsilon=-0.54$, after the flow was established, the calculation was continued with new boundary conditions which signified the departure of the body from the rarefied region and its return to air at normal density. This setup of the problem simulates the interaction of the body with a rarefied volume of finite length, with linear dimensions of the order of the length L of the cone.

The calculations yielded a detailed spatial–temporal picture of the flow as the body enters and leaves the rarefied region. We present some figures illustrating these results. Figure 2 gives an idea of the density distribution in the shock layer during the interaction of a cone and a rarefied volume with $\varepsilon=+0.54$. The distribution corresponds to the time $t=0.56$, when the cone has entered the rarefied region to a depth of $\sim 1/2$ its length. The figure shows the meridional cross section of the plane passing through the direction $\varphi=0-180°$. The magnitude of the density is proportional to the intensity of the shading. The dashed curve, as in Figs. 4 and 5, indicates the boundary of the rarefied region. It is

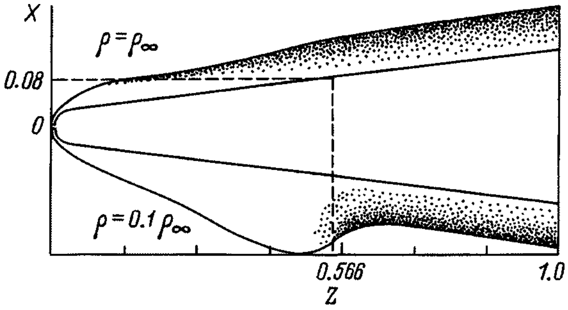

FIG. 2. Density of air in the shock layer in the meridional cross section $\varphi = 0-180°$.

quite clear how the shock surface is distorted in the low-density region. As the body enters the rarefied volume, the part of the shock layer moving through the rarefied area is ''puffed up.'' According to the initial conditions for the problem, the speed of sound in the rarefied air is a factor of three higher than in the dense air, while the density is an order of magnitude lower. This leads to a reduction in the intensity P_W/P_∞ of the bow shock in the rarefied air and to expansion of the shock layer (P_W is the pressure in the bow shock). To complete the picture, Fig. 3 shows the three-dimensional shape of the surface of the bow shock for the same interaction conditions $\varepsilon = +0.54$ and at the same time $t = 0.56$ as in Fig. 2. The spatial deformation of the initially conical shape of the bow shock as the body enters the rarefied volume is quite evident in Fig. 3.

We now give some examples of the distributions of the gasdynamic parameters over the angular and radial coordinates. To do this, we examine the flow parameters in planes perpendicular to the axis of the cone. According to the law of ''plane cross sections,''[11] during the hypersonic flight at velocity U_∞ of a thin circular cone with a half angle β, the distribution of the gas parameters in planes perpendicular to the direction of motion of the cone is equivalent to the one-dimensional flow which arises during expansion of a circular cylinder into a gas at rest with its radius increasing as $R = U_\infty \tan\beta t$. For the motion of a blunt cone, the analogy is retained with a correction for the initial momentum and energy imparted by the blunt end to the motionless gas. Figure 4 gives an idea of the pressure distribution in the $Z = 0.55$ cross section at time $t = 0.88$ as the cone moves in a rarefied volume with an impact parameter $\varepsilon = +0.54$. This distribution can be regarded as the result of the expansion of a cylindrical piston in a motionless gas with a nonuniform density. Recall that, according to the statement of the problem, the density ρ_0 of the air in the rarefied region is an order of magnitude lower than the initial density ρ_∞. This means that the pressure at that part of the cone which moves in the dense gas is higher than the pressure on the surface immersed in the rarefied air. At the same time, the radial propagation velocity of the shock in the dense air is lower than in the rarefied air. At the boundary of the regions of different density, where shock waves with different intensities and different velocities are joined, a three-wave configuration develops. This figure also illustrates the shape of the surface of the shock wave at this cross section, which is characterized by substantial ''puffing out'' of the shock layer moving through the rarefied air. The excess pressure $\Delta P = P_T - P_\infty$ on the conical surface in the rarefied region is roughly a factor of 8 smaller than ΔP in the dense region of the flow, which is close to the expected value, ~ 10, given by the formula $\Delta P \sim \rho_\infty U_\infty^2 \sin^2\beta$.

In Fig. 5 we show a fragment of the velocity distribution at the same cross section $Z = 0.55$ and the same time $t = 0.88$. The circular motion of the gas in the region where air with different densities comes into contact is also evident

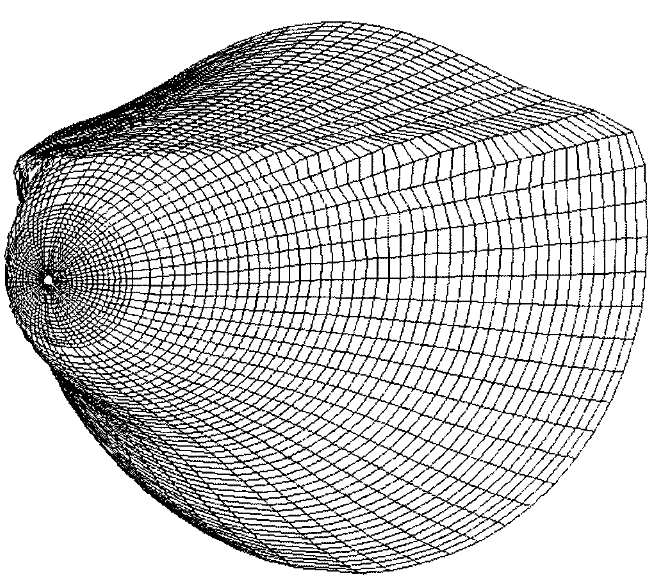

FIG. 3. The shape of the bow shock wave of a blunt cone as it interacts with a rarefied volume.

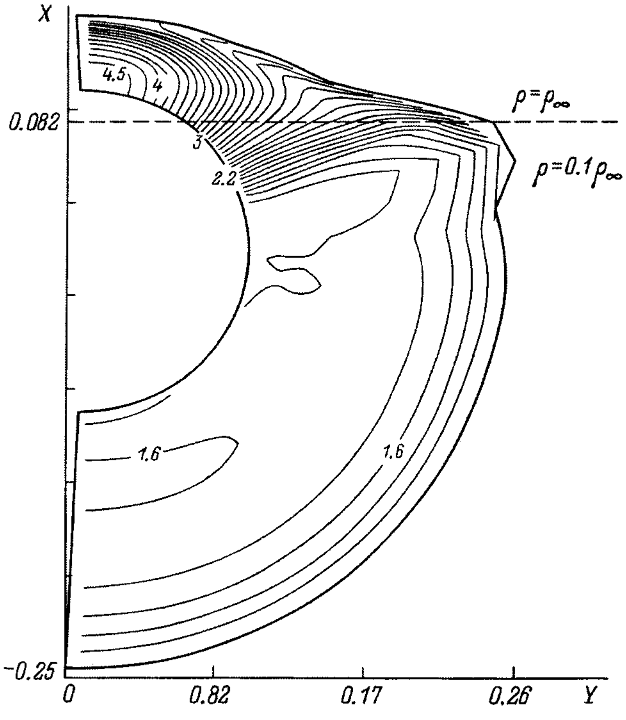

FIG. 4. Constant-pressure contours in the shock layer in the cross section $Z = 0.55$.

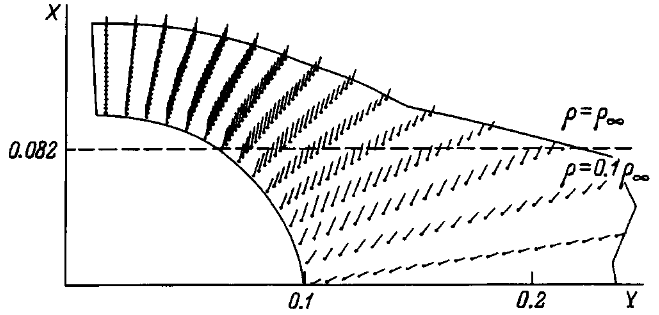

FIG. 5. The velocity field in the shock layer at the cross section $Z=0.55$ for angles $0<\varphi<90°$ (view from the side of the nose).

there. The pressure in the rarefied region is greater after passage of the shock than the pressure in the unperturbed air. Air from the rarefied region expands laterally into the surrounding unperturbed gas, forming a transition region where shocks with different propagation velocities meet. In ending this discussion of the three-dimensional flow, we should mention that a series of three-dimensional calculations of the movement of a cone along the boundary of two regions with densities differing by 10 or more times showed that the transition from the dense to the rarefied flow at the surface of the cone takes place within a narrow region near the contact boundary and is not accompanied by significant mass transfer between the dense and less dense regions of the flow.

AERODYNAMIC CHARACTERISTICS

As we have said, the change in the aerodynamic characteristics of the body as it interacts with the rarefied region is of considerable interest for applications. In our three-dimensional calculations, we evaluated the aerodynamic coefficients of normal force C_n, drag force C_τ, and pitching moment C_m of the cone. These coefficients provide a most complete characterization of the aerodynamic loads experienced by the body as it enters and moves in the rarefied region. In our case, C_n is defined as the ratio of the projection of the aerodynamic force F_X along the X axis to the quantity $0.5\rho_\infty U_\infty^2 S$, C_τ equals the ratio of F_Z to the same quantity, and C_m is the ratio of the pitching moment M, calculated with respect to the center of mass of the body, to $0.5\rho_\infty U_\infty^2 SL$, where S is the area of the midsection of the body.

Figure 6 shows the time variation in the coefficients of normal force and pitching moment for the same set of calculations in which we examined the entry and exit of the body from the rarefied region. Essentially, this calculation models the interaction of the cone with a rarefied volume of length $\Delta\approx 2L$. An impact parameter of $\varepsilon=-0.54$ means that during stationary motion of the body along the boundary of the regions of different density, $\sim$20% of the surface area of the body is immersed in the rarefied air. We now explain the time dependences $C_n(t)$ and $C_m(t)$. For uniform flow around the cone, C_n and C_m are zero. At time $t=0$ the bow shock comes into contact with the rarefied channel. As the cone enters the rarefied volume, the area of the rarefied region on the surface of the cone grows, so that the lateral force F_n and pitching moment M increase. These quantities reach their stationary values over a time $t\sim L/U_\infty$, which corresponds to the body's entering the rarefied air along its entire length. At $t\sim 1.5$, the cone begins to return to the normal atmosphere. The coefficients C_m and C_n then decrease, since there is a return to a symmetric flow around the body, go to zero when the body is again moving in homogeneous air. Note the nonmonotonic character of $C_m(t)$ owing to the sensitivity of the magnitude and sign of the pitching moment to the size of the rarefied region and its position on the surface of the cone.. Evidently, the cause of the lateral force and rotational moment is the rarefied region which develops at the surface of the body as it moves through the low-density air. How long this aerodynamic load acts is determined by the existence time of the rarefied region.

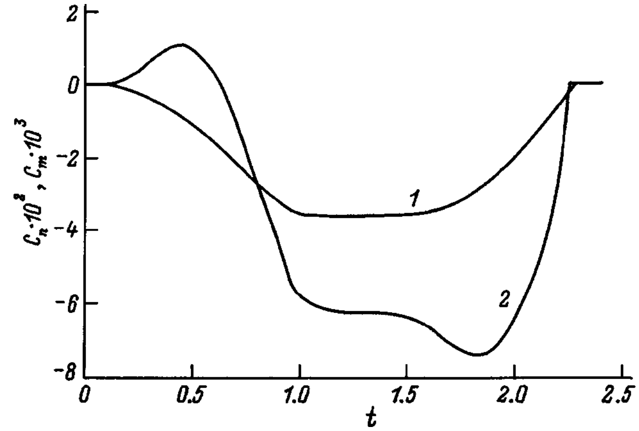

FIG. 6. The aerodynamic coefficients C_n (*1*) and C_m (*2*) as functions of time during the interaction of a cone with a rarefied region.

Let us compare the computational results obtained with the same formulation of the problem, but with different degrees ε of immersion of the body in the rarefied air. Figures 7 and 8 show how the normal force C_n, drag C_τ, and pitching moment C_m coefficients of the cone depend on the degree of immersion of the body in rarefied air in the case of stationary motion. Besides the three-dimensional calculations, which are indicated by the data points, the smooth

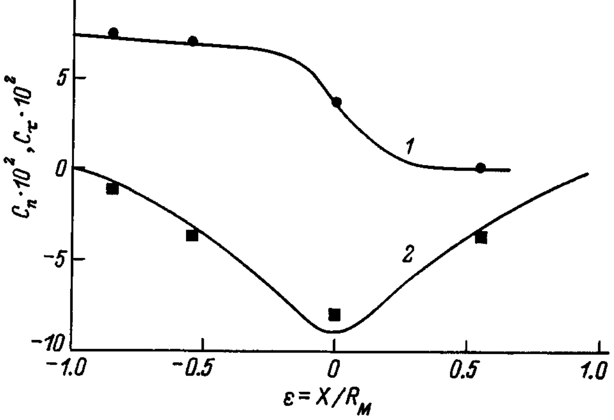

FIG. 7. The drag (*1*) and lateral force (*2*) coefficients as functions of the degree of immersion of the body in a rarefied volume.

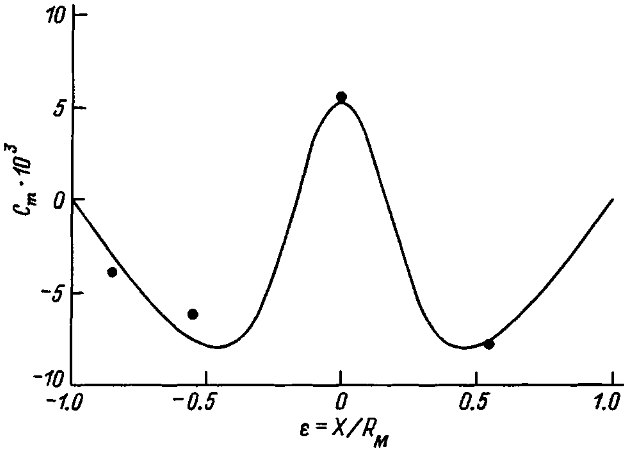

FIG. 8. The pitch coefficients as functions of the degree of immersion of the body in a rarefied volume.

curves in Figs. 7 and 8 show the dependences obtained from two-dimensional calculations of a uniform flow around the body. These coefficients $C_{\tau 2}$, C_{n2}, and C_{m2} were calculated using an approximation for the motion of a body along the interface between two regions with different densities. In this approach, immersion in rarefied air was modeled by specifying zero pressure on that part of the body's surface lying in the rarefied volume while keeping the pressure on the rest of the body unchanged. Here and in the following, the subscript 2 denotes coefficients obtained in this approximation for two dimensions. The observed agreement between the approximate and exact calculations of the interaction shows that the aerodynamic coefficients can be calculated in terms of the two-dimensional approximation with an accuracy that is acceptable for practical purposes. The physical reason for the good agreement between the three-dimensional calculations and the approximate calculations is that, as stated above, in high-speed supersonic flows with a tangential shock, the low-pressure region at the surface of the body remains localized, despite the three-dimensional character of the flow.

We now comment briefly on the variation of the aerodynamic coefficients with ε, using Figs. 7 and 8. C_n and C_τ behave in fairly obvious fashion with increasing ε or, equivalently, with increasing surface area immersed in the low-density air. The lateral force F_n initially rises monotonically, reaches a maximum $F_{n\max}$ when half of the body's surface and of the bow shock are in the rarefied air, and then begins to decrease. The drag force F_τ decreases monotonically as the degree of immersion increases. Note that the maximum drag force $F_{\tau\max}$ is less than the maximum lateral force $F_{n\max}$.

The behavior of the pitching moment M relative to the center of mass of the body and the change in its sign, which determines the direction of rotation of the cone, are of interest. First, Fig. 8 shows that the pitching moment is very sensitive to the magnitude of ε, i.e., to the degree of immersion in the rarefied air. It is clear that $C_{m2}(\varepsilon)$ varies rapidly over the interval $-1 \leqslant \varepsilon \leqslant 1$, changing sign twice. Second, the $C_{m2}(\varepsilon)$ curve has two minima, which are confirmed by the three-dimensional calculations. They occur at $\varepsilon \sim \pm 0.5$, i.e., when roughly 1/5 or 4/5 of the body's surface area is immersed in the rarefied air. Here the moment is directed so that the bottom part of the body will be immersed in the rarefied air. When $\varepsilon = 0$ and, in general, for $|\varepsilon| \leqslant 0.16$, the nose of the cone is immersed in rarefied air. The high sensitivity of the magnitude and sign of the moment to the position of the contact boundary makes it difficult to predict the possible motion of the cone when ε is close to zero. However, for $|\varepsilon| > 0.5$, i.e., for almost complete or, in the opposite case, for negligible immersion of the object in the rarefied volume, the results of the interaction are entirely predictable.

CONCLUSION

The cycle of two and three-dimensional calculations performed here has revealed a mutual position of an extended rarefied region and a blunt cone for which the aerodynamic coefficients of a cone flying along a boundary with a rarefied region are maximal. Thus, the maximum pitching moment M_{max} occurs when either $\sim 20\%$ or $\sim 80\%$ of the lateral surface area of the cone is immersed in the rarefied volume. The cone experiences a maximum lateral force $F_{n\max}$ when roughly half the body's surface is immersed in the rarefied air. Finally, when $>60\%$ of the body's area is immersed in the rarefied region, the drag force F_τ becomes negligible.

[1)] Here we note some errors in Ref. 7. In the last term of the last equation of the system of Eqs. (1) of that paper, the r should be deleted. The total energy per unit volume is $E = \rho(\varepsilon + (U^2 + V^2 + W^2)/2)$.

---

[1] R. Ya. Tugazakov, Uch. Zap. Tsentral. Aéro-Gid. Inst. **11**, 139 (1980).
[2] V. I. Artem'ev, V. I. Bergel'son, I. V. Nemchinov *et al.*, Izv. AN SSSR. MZhG, No. 5, 146 (1989).
[3] V. P. Goloviznin and I. V. Krasovskaya, Zh. Tekh. Fiz. **61**(12), 12 (1991) [Tech. Phys. **36**, 1332 (1991)].
[4] O. M. Velichko, V. D. Urlin, and B. P. Yakutov, Vopr. At. Nauk Tekh. Ser. Teor. Prikl. Fiz., No. 1, pp. 20–22 (1992).
[5] P. Yu. Georgievskiĭ and V. A. Levin, Izv. Ross. Akad. Nauk Mekh. Zhidk. Gazov, No. 4, 174 (1993).
[6] V. Yu. Borzov, I. V. Rybka, and A. S. Yur'ev, Inzh.-Fiz. Zh. **67**, 355 (1994).
[7] O. M. Velichko, V. D. Urlin, and B. P. Yakutov, Zh. Tekh. Fiz. **65**(5), 31 (1995) [Tech. Phys. **40**, 422 (1995)].
[8] S. K. Godunov, A. V. Zabrodin, M. Ya. Ivanov *et al.*, *Numerical Solution of Multidimensional Gasdynamics Problems* [in Russian], Nauka, Moscow (1976), 400 pp.
[9] O. M. Velichko and A. N. Razin, Vopr. At. Nauk Tekh. Ser. Mat. Model. Fiz. Protsess., No. 3, pp. 3–6 (1991).
[10] O. M. Velichko and T. I. Kravchenko, *ibid*, pp. 71–75.
[11] G. G. Chernyĭ, *Gas Flows at High Supersonic Velocities* [in Russian], Fizmatgiz, Moscow (1959), 220 pp.

Translated by D. H. McNeill

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Determination of model parameters for the hydrogen permeability of metals

Yu. V. Zaika

Petrozavodsk State University, 185640 Petrozavodsk, Russia

(Submitted July 3, 1997)
Zh. Tekh. Fiz. **68**, 38–42 (November 1998)

The nonlinear inverse problem of determining the permeability parameters of metals for hydrogen from experimental data is examined. The model includes adsorption–desorption processes as well as diffusion. An algorithm is proposed for determining the model parameters from a known desorption flux without the need for writing specialized computer programs. © *1998 American Institute of Physics.* [S1063-7842(98)00711-9]

INTRODUCTION

Interest in the interaction of hydrogen and its isotopes with metals exists on many levels.[1–6] It is sufficient to mention problems in power production, protection of construction materials from hydrogen corrosion, chemical reactor design, the building of rockets, and vacuum technology. Physical and chemical phenomena on the surface, as well as diffusion processes inside metals, play an important role (Ref. 3, pp. 177–206). The transport parameters for hydrogen also depend on the technical peculiarities of the process by which a specific batch of metal has been obtained, as well as on the surface treatment. This limits the use of various tabulated data. Crude estimates of the parameters (with errors within a few orders of magnitude) are usually available. There is a need for an algorithm which can be used to refine these values on the basis of experimental data from specific materials.

An iterative numerical algorithm has been proposed for determining hydrogen transport models.[7] In this paper it is shown that if the interaction with traps is assumed to be a small perturbation and the aim is only to determine the main parameters (in the framework of the chosen model), then the problem can be solved without the experimenter's having to write a specialized computer program.

MATHEMATICAL MODEL

The experimental technique for studying permeation involves creating a rather high constant pressure $p_0(t)=\overline{p_0}$ of gaseous hydrogen as a discontinuity on the inlet side of an initially dehydrogenated membrane (vacuum vessel barrier) that has been heated to a fixed temperature $T(t)=\overline{T}$. On the outlet side, the gas is continuously removed by a vacuum system. We treat the exiting desorption flux of hydrogen as the experimental data. For concreteness, the author takes the experimental apparatus of Gabis *et al.*[8] as an example.

We take the following mathematical model (Ref. 3, pp. 177–206):

$$\frac{\partial c}{\partial t}=D(T)\frac{\partial^2 c}{\partial x^2},\quad (t,x)\in Q_{t^+}=(0,t^+)\times(0,l), \tag{1}$$

$$c(0,x)=\varphi(x),\quad x\in[0,l], \tag{2}$$

$$c_0(t)=c(t,0)=g(T)q_0(t),$$

$$c_l(t)=c(t,l)=g(T)q_l(t), \tag{3}$$

$$\frac{d}{dt}q_0(t)=\mu s(T)p_0(t)-b(T)q_0^2(t)+D(T)\frac{\partial c}{\partial x}(t,0), \tag{4}$$

$$\frac{d}{dt}q_l(t)=-b(T)q_l^2(t)-D(T)\frac{\partial c}{\partial x}(t,l),\quad t\in[0,t^+], \tag{5}$$

$$g(T(0))q_0(0)=c_0(0)=\varphi(0),$$

$$g(T(0))q_l(0)=c_l(0)=\varphi(l). \tag{6}$$

Here $c(t,x)$ is the concentration of diffusing (atomic) hydrogen, $q_0(t)$ and $q_l(t)$ are the surface concentrations ($x=0,l$), $D(T)$ is the diffusion coefficient, $g(T)$ is the matching coefficient between the concentrations on the surface and in the volume near the surface of the membrane, μ is a kinetic constant, $s(T)$ is the sticking coefficient of the surface for hydrogen in the gaseous phase, and $b(T)$ is the desorption coefficient. If the membrane is dehydrogenated at the initial time $t=0$, then $\varphi(x)=0$. Equations (4) and (5) are the flux balance equations. The desorption flux is modeled by a quadratic variation. For other gases a different functional dependence can be used; this is not of fundamental importance below. The last terms on the right-hand sides of Eqs. (4) and (5) correspond to an efflux or influx of hydrogen atoms to the surface owing to diffusion within the membrane volume. In Eq. (5) there is no term $\mu s(T)p_l(t)$, since for a sufficiently fast vacuum system the pressure $p_l(t)$ at the outlet is very low and there is negligible return of hydrogen desorbed from the outlet surface back to the surface. The initial and boundary conditions are matched in the sense of Eq. (6).

It is required to determine $D(T)$, $g(T)$, $s(T)$ and $b(T)$ for a specific material from the outlet desorption flux

$$J(t)=b(T)q_l^2(t)=b(T)g^{-2}(T)c_l^2(t),$$

$$T=T(t),\quad t\in[0,t_*]. \tag{7}$$

We shall omit the word ''flux,'' assuming that the surface has unit area. The time t_* at which the experiment ends is determined by the approach to a stationary flux $J(t)\approx\overline{J}=\text{const}$, $t\geqslant t_*$ $(t_*<t^+)$.

The Arrhenius law

$$D=D_0\exp(-E_D/[RT(t)]),\ldots,$$

$$b=b_0\exp(-E_b/[RT(t)])$$

is usually applied to hydrogen.

Other models for the temperature dependence are possible. In the following we discuss an algorithm for determining D, g, s, and b for $T(t)=\overline{T}=\text{const}$. This is the basic problem, a nonlinear inverse problem of mathematical physics. Information on the values of D, g, s, and b at different temperatures makes it possible to determine the parameters in $D(T),\ldots,b(T)$, as well (in the Arrhenius case, D_0, $E_D,\ldots,b_0$, and E_b). The real noisiness of the measurements is taken into account in constructing the algorithms: in the final formulas the experimental data enter in the form of an integral over $[0,t_*]$, which ensures a noise-free determination. The model (1)–(6) has its limits of applicability. Thus, the idea of the algorithm is laid out in some detail and permits some variation in the equations.

The boundary conditions (3)–(6) are nonstandard, so a few words are necessary on the mathematical justification of the model. If $\varphi\in H^1(0,l)$ and $T\in C^1[0,t^+]$, then, with the coefficients restricted in accordance with their physical meaning, there is a unique solution $c(t,x)\in H^{1,2}(Q_{t^+})$. The function $c(t,x)$ in Q_{t^+} satisfies Eq. (1), is uniformly continuous in the rectangular region Q_{t^+}, and is continued continuously to the closure $\overline{Q}_{t^+}=[0,t^+]\times[0,l]$. Equation (2) is satisfied for $x\in[0,l]$. The gradients $c_x(t,0)$ and $c_x(t,l)$ are defined over $c\in H^{1,2}$ as elements of $L_2(0,t^+)$. After substituting them in Eqs. (4) and (5), we have ordinary differential equations with the initial data (6). Their solutions $q_0,q_l\in H^1(0,t^+)$ satisfy Eqs. (4) and (5) in the sense of hydrogen permeation parameters over $[0,t^+]$. Finally, after substituting $q_0(t)$ and $q_l(t)$ in Eq. (3), we obtain an identity over $t\in[0,t^+]$. If we take $c(t,\cdot)\in H^1(0,l)$, $t\geqslant 0$ for the phase state, then the model (1)–(6) is a nontrivial example of a nonlinear semidynamical system in the Hilbert space $H^1(0,l)$. The study of this system is also of mathematical interest.[9] As the smoothness of the initial data for the boundary value problem (1)–(6) increases, the smoothness of the solution $c(t,x)$ will also increase.

THE DETERMINATION ALGORITHM

Let $T(t)=\overline{T}$, $p_0(t)=\overline{p}_0$, and $\varphi(x)$ be fixed. Measurements of $J(t)=bq_l^2(t)$ are directly related to surface processes. Thus, it is appropriate to ''exclude'' the diffusion equation in the membrane volume. Here the appropriate mathematical apparatus is integration by parts. This technique leads to so-called conjugate equations.[10]

For an arbitrary function $\psi(t,x)$ that is sufficiently smooth in $\overline{Q}_{t*}$, in view of Eq. (1) we have

$$\begin{aligned}0&=\int_0^{t_*}\int_0^l\psi(t,x)(c_t-Dc_{xx})dxdt\\&=\int_0^l\{\psi(t_*,x)c(t_*,x)-\psi(0,x)c(0,x)\}dx\\&\quad-D\int_0^{t_*}\{\psi(t,l)c_x(t,l)-\psi(t,0)c_x(t,0)\}dt\\&\quad+D\int_0^{t_*}\{\psi_x(t,l)c(t,l)-\psi_x(t,0)c(t,0)\}dt\\&\quad-\int_0^{t_*}\int_0^l c(t,x)(\psi_t+D\psi_{xx})dx/,dt.\end{aligned}\tag{8}$$

It is easy to make the last, ''volume'' term go to zero by the choice of $\psi(t,x)$. The remaining terms are related to the boundary conditions. We shall analyze (to within the values of the *a priori* unknown model parameters) the information content of the pair $p_0(t),J(t)$. The functions $c_l(t)$ and $c_x(t,l)$ can be expressed in terms of $J(t)$ by virtue of Eqs. (3), (5), and (7). But knowledge of only $p_0(t)=\overline{p}_0$ at the inlet is not very informative: only the term on the right-hand side is known in the differential equation (4). This is not a mathematical deficiency: in order to correctly identify a ''black box'' it is necessary to know the input and output. It might be possible to determine the flux $Dc_x(t,0)$ from the hydrogen feed rate, but the high background $\overline{p}_0$ interferes. Measuring the concentration $c_0(t)(q_0(t))$ is also problematical. It is an entirely different matter to measure $J(t)$ under vacuum pumping conditions.

There is, nevertheless, a way out of this situation; it is only necessary to take into account the difference in the rates of surface processes ($x=0$, $x=l$) across the enormous pressure drop (7–9 orders of magnitude).

Let $\psi(t,x)$ satisfy the equation conjugate to Eq. (1) with only one boundary condition,

$$\frac{\partial\psi}{\partial t}=-D\frac{\partial^2\psi}{\partial x^2},\quad (t,x)\in Q_{t_*},\tag{9}$$

$$\psi(t,0)=0,\quad t\in[0,t_*].\tag{10}$$

Under condition (9) the last term in Eq. (8) will vanish, and by virtue of Eq. (10), the flux $Dc_x(t,0)$, whose values are not known from the experimental setup, does not appear in Eq. (8). The solution of Eqs. (9) and (10) is easily found by separation of variables: $\psi(t,x)=\beta(t)\gamma(x)$, with $\gamma(0)=0$. There are infinitely many such solutions, a fact which is important in the subsequent discussion. Equation (8) now takes the form

$$\int_0^l\psi c\Big|_{t=0}^{t_*}dx-D\int_0^{t_*}\psi c_x\Big|_{x=l}dt+D\int_0^{t_*}\psi_x c\Big|_{x=0}^{l}dt=0.\tag{11}$$

The further strategy in using Eq. (11) is as follows: we choose any solution $\psi=\psi_i(t,x)$ of Eqs. (9) and (10) and substitute in Eq. (11) the expressions for $c(0,x)$, $c(t_*,x)$, $c_x(t,l)$, and $c_{0,l}(t)$ in terms of the known information and the parameters D, g, s, and b. As a result, we obtain an equation $f_i(D,G,s,b)=0$. One must take a number of such equations sufficient for determining D, g, s, and b.

The function $c(0,x)=\varphi(x)$ is specified by the initial conditions $(\varphi(x)=0)$. Since $\overline{p}_0=\text{const}$, after some time a stationary output flux $J(t)=\overline{J}=\text{const}$, $t\geqslant t_*$, is established. For $t\geqslant t_*$ all the time derivatives in Eqs. (1)–(6) can be assumed to be zero, so we obtain a linear stationary concentration distribution $c(t,x)=c(t_*,x)$, $t\geqslant t_*$, with $c(t_*,x)=\xi_1\cdot(x-l)+\xi_2$. The model of Eqs. (1)–(6) reflects these simple experimental facts (verified numerically). Let us calculate ξ_1 and ξ_2. From Eq. (7), we have

$$\xi_2=c(t_*,l)=gq_l(t_*)=gb^{-1/2}\overline{J}^{1/2}.$$

We find the slope ξ_1 of the straight line $c(t_*,x)$ from Eq. (5):

$$\xi_1=c_x(t,l)=-D^{-1}\overline{J}\ (\dot{q}_l=0,\quad t\geqslant t_*).$$

Finally, for $t\geqslant t_*$, we obtain

$$c(t,x)=c(t_*,x)=D^{-1}\overline{J}\cdot(l-x)+gb^{-1/2}\overline{J}^{1/2}. \tag{12}$$

Thus, with time, at the outlet $(x=l)$ a concentration $\overline{c}_l$ proportional to $\overline{J}^{1/2}$ develops, while the diffusion flux $Dc_x(t,x)$ does not vary over the thickness of the membrane and has an absolute value equal to the desorption flux $\overline{J}$.

Now, using Eq. (5), we transform the second integral in Eq. (11),

$$\begin{aligned}
-&\int_0^{t_*}\psi(t,l)Dc_x(t,l)\,dt\\
&=\int_0^{t_*}\psi(t,l)(\dot{q}_l(t)+J(t))\,dt=\psi(t,l)q_l(t)\big|_0^{t_*}\\
&-\int_0^{t_*}\dot{\psi}(t,l)q_l(t)\,dt+\int_0^{t_*}\psi(t,l)J(t)\,dt\\
&=\psi(t_*,l)b^{-1/2}\overline{J}^{1/2}+\int_0^{t_*}\{\psi(t,l)J(t)\\
&-\dot{\psi}(t,l)b^{-1/2}J^{1/2}(t)\}\,dt,\\
&(q_l(0)=0,\quad q_l(t)=b^{-1/2}J^{1/2}(t),\\
&q_l(t_*)=b^{-1/2}\overline{J}^{1/2}).
\end{aligned} \tag{13}$$

The last expression in Eq. (13) already contains known quantities and the parameter b.

The last integral in Eq. (11) still has to be specified. The concentration $c_l(t)$ is expressed in terms of $J(t)$ and the parameters through Eq. (7): $c_l(t)=gb^{-1/2}J^{1/2}(t)$. How do we calculate the integral of $\psi_x(t,0)c_0(t)$ with sufficient accuracy, given that $c_0(t)$ is not accessible to measurement? At the outlet, because of the vacuum pumping, the surface is depleted of hydrogen, and its buildup is a limiting factor (Eq. (5)). At the input, with $p_0\gg p_l$, the surface is rapidly saturated to a level $\overline{q}_0$ corresponding to $\overline{p}_0$, with a subsequent relatively slow leak-off of the diffusant into the bulk. The duration of the transient process is very short compared to the time t_* for approach to the stationary level of $\overline{J}$. In an experiment, this condition for the concentration jump at the inlet can be realized to good accuracy: one increases $\overline{p}_0$ and the thickness l as necessary. Then, in calculating the integral of $\psi_x(t,0)c_0(t)$ over the segment $[0,t_*]$ we can set $c_0(t)\approx\overline{c}_0=g\overline{q}_0$. The concentration $\overline{c}_0$ is determined by Eq. (12) $(x=0)$. For the model (1)–(6) these qualitative considerations are confirmed by numerical simulations.

We now have everything necessary to set up the specific equations $f(D,g,s,b)=0$. First, let us extract the maximum possible from the mapping $\overline{p}_0\rightarrow\overline{J}$. If the inlet concentration is steady, then Eq. (4) implies that

$$\dot{q}_0=0\rightarrow\overline{c}_0=gb^{-1/2}(\mu s\overline{p}_0+Dc_x(t,0))^{1/2}$$

$$(t\geqslant\varepsilon,\ \varepsilon\ll t_*).$$

The flux $Dc_x(t,0)$ does not vary $(t\geqslant\varepsilon)$, and it has already been calculated for $t\geqslant t_*$: $Dc_x(t,0)=Dc_x(t,l)=-\overline{J}$, $t\geqslant t_*$. Thus,

$$\overline{c}_0=gb^{-1/2}\sqrt{\mu s\overline{p}_0-\overline{J}}. \tag{14}$$

The quantity under the radical is positive. The meaning of Eq. (14) $(\overline{J}=\mu s\overline{p}_0-bg^{-2}\overline{c}_0^{\,2})$ is the following: after saturation at the inlet to a value $\overline{c}_0$ corresponding to $\overline{p}_0$, a dynamic equilibrium is established. The permeating flux equals the difference between the flux $\mu s\overline{p}_0$ incident on the surface and the desorption flux back into the chamber volume. At the outlet, on the other hand, the level $\overline{J}$ is reached only at the time t_*. Comparing Eqs. (14) and (12) $(x=0)$, we obtain the first equation $f_1=0$,

$$gb^{-1/2}(\mu s\overline{p}_0-\overline{J})^{1/2}-gb^{-1/2}\overline{J}^{1/2}-lD^{-1}\overline{J}=0. \tag{15}$$

Equation (15) can be used to find s and the combination $X=Dgb^{-1/2}/l$. To do this, we modify the permeation experiment. Initially, as described above, with an injection pressure $\overline{p}_{01}$ and $\varphi(x)=0$, we wait until the time t_* when $J=\overline{J}_1$ is established. Then, we raise the pressure suddenly to $\overline{p}_{02}>\overline{p}_{01}$ and wait a further time Δt_* until $\overline{J}_2$ is established. We substitute the two pairs $\overline{p}_{0i}$, $\overline{J}_i$ in Eq. (15) and shift the last term to the right. Then

$$\overline{J}_1/\overline{J}_2=[(\mu s\overline{p}_{01}-\overline{J}_1)^{1/2}-\overline{J}_1^{1/2}][(\mu s\overline{p}_{02}-\overline{J}_2)^{1/2}-\overline{J}_2^{1/2}]^{-1}.$$

Writing

$$y=(\mu s\overline{p}_{01}-\overline{J}_1)^{1/2},\quad d_1=\overline{J}_1^{1/2}(1-\overline{J}_1^{1/2}/\overline{J}_2^{1/2}),$$

$$d_2=\overline{p}_{02}\overline{J}_1^2/(\overline{p}_{01}\overline{J}_2^2),$$

we obtain $\mu s\overline{p}_{02}=(y^2+\overline{J}_1)\overline{p}_{02}/\overline{p}_{01}$ and

$$y-d_1=\{(y^2+\overline{J}_1)d_2-\overline{J}_1^2/\overline{J}_2\}^{1/2}. \tag{16}$$

Squaring Eq. (16) gives

$$(1-d_2)y^2-2d_1y+d_1^2+d_3=0,\quad d_3=\overline{J}_1(\overline{J}_1/\overline{J}_2-d_2). \tag{17}$$

By virtue of the fact that $\overline{p}_{02}>\overline{p}_{01}\Rightarrow\overline{J}_2>\overline{J}_1$ and Eq. (15), we have $d_1>0$, $d_2<1$, and $d_3<0$, and the quadratic equation (17) has two real roots, with $y_1>y_2$. The larger root y_1 has physical significance. It is greater than d_1 and is consistent with Eq. (16). The second root, $y_2<d_1$ is inconsistent with Eq. (16) and appears because the square was taken. From the value $y=y_1$ we find s and, with Eq. (15), the combination $X=Dgb^{-1/2}/l$.

As we see, an analysis of just the stationary values cannot be used to determine D, g, and b, but only gives the combination X. We determine some additional equations from the transient process, using Eqs. (11)–(14).We take the simplest solution of Eqs. (9) and (10), viz., $\psi(t,x)=x/l$. For $\psi=x/l$, Eq. (11) with Eq. (13) simplifies to

$$l^{-1}\int_0^l xc(t_*,x)dx+b^{-1/2}\overline{J}^{1/2}+\int_0^{t_*}J(t)dt$$
$$+Dl^{-1}\int_0^{t_*}c\Big|_{x=0}^{l}dt=0.$$

For computational reasons it is convenient to change to the variables $x_1=l^2/D$, $x_2=lg$, and $x_3=b^{-1/2}$. After substituting $c(t_*,c)$ from Eq. (12), $c_l(t)=gb^{-1/2}J^{1/2}(t)$ and $c_0(t)\approx\overline{c}_0$, from Eq. (12) for $x=0$ (or from Eq. (14)), we obtain

$$f_2=x_1\overline{J}_1/6+x_1X\overline{J}_1^{1/2}/2+x_3\overline{J}_1^{1/2}+A_1=0, \tag{18}$$

where

$$A_1=S_1+XS_{1/2}-\overline{J}_1t_*-X\overline{J}_1^{1/2}t_*,\quad S_\sigma=\int_0^{t_*}J^\sigma(t)dt$$

$$(X=Dgb^{-1/2}/l=x_1^{-1}x_2x_3\rightarrow x_2x_3=x_1X).$$

We use the same $\psi(t,x)=x/l$ in the time interval $[t_*,t_*+\Delta t_*]$. The calculations are the same. It should only be noted that when the time origin is shifted to t_* in Eq. (11), we have $\psi(x)\neq 0$ $(\varphi(x)=c(t_*,x))$ and in Eq. (13), $q_l(0)\neq 0$ $(q_l(0)=b^{-1/2}\overline{J}_1^{1/2})$, and we get

$$f_3=-x_1\overline{J}_1/6-x_1X\overline{J}_1^{1/2}/2-x_3\overline{J}_1^{1/2}+x_1\overline{J}_2/6$$
$$+x_1X\overline{J}_2^{1/2}/2+x_3\overline{J}_2^{1/2}+\Delta A_1=0,$$
$$\Delta A_1=\Delta S_1+X\Delta S_{1/2}-\overline{J}_2\Delta t_*-X\overline{J}_2^{1/2}\Delta t_*,$$
$$\Delta S_\sigma=\int_{t_*}^{t_0}J^\sigma(t)dt,\quad t_0=t_*+\Delta t_*. \tag{19}$$

Equations (18) and (19) are a system of two linear algebraic equations in x_1 and x_3. Eliminating the variable x_3, we obtain

$$f_2-\xi f_3=0,$$
$$\xi=\overline{J}_1^{1/2}/(\overline{J}_2^{1/2}-\overline{J}_2^{1/2})\rightarrow x_1=6(A_1-\xi\Delta A_1)/(\overline{J}_1\overline{J}_2)^{1/2}. \tag{20}$$

After substituting x_1 in Eq. (18) (or (19)), we find x_3 and then $x_2=Xx_1/x_3$.

For convenience of application, we give the sequence of solving the problem of model determination:

1. Provisional experimental scheme: $T=\overline{T}$, $t=0\rightarrow(\varphi(x)=0,p=\overline{p}_{01})$, $t=t_*\rightarrow(J=\overline{J}_1,\varphi(x)=c(t_*,x),p=\overline{p}_{02}>\overline{p}_{01})$, $t=t_*+\Delta t_*\rightarrow J=\overline{J}_2$.

2. From $\overline{p}_{0i}$ and $\overline{J}_i$ we calculate d_1, d_2, d_3 and the larger root y_1 of the quadratic equation (17), we determine

$$s=(y_1^2+\overline{J}_1)/(\mu\overline{p}_{01}),$$

and we write Eq. (15) in the form

$$X_i=\overline{J}_1/[(\mu s\overline{p}_{0i}-\overline{J}_i)^{1/2}-\overline{J}_i^{1/2}],\quad X=(X_1+X_2)/2.$$

3. Using the quadrature formula, we calculate the integrals S_1, $S_{1/2}$, ΔS_1, and $\Delta S_{1/2}$, and then the quantities A_1 and ΔA_1 (see Eqs. (18) and (19)).

4. With Eq. (20) we find x_1 $(D=l^2/x_1)$, then $x_3=x_{31}$ and $x_3=x_{32}$ from Eqs. (18) and (19) and, finally, $x_3=(x_{31}+x_{32})/2$ $(b=x_3^{-2})$, $x_2=Xx_1/x_3$ $(g=x_2/l)$.

In the numerical experiments confirming the efficiency of the algorithm, it was assumed that $l=0.02$ cm and $\mu=1.46\times 10^{21}$ mol/cm$^2\cdot$s$\cdot$Torr, and the following reference values were varied: $D=10^{-6}$ cm^2/s, $g=10$ cm^{-1}, $b=10^{-17}$ cm^2/s, $s=10^{-4}$, and $\overline{p}_0=0.1$ Torr.

We briefly examine the possibility of using other ψ. If ψ is restricted only by Eq. (9), then in Eq. (8) we can take $\psi=(x-l)/l$. This eliminates the last term and the flux $Dc_x(t,l)$ from the formula. The term containing $Dc_x(t,0)$ has to be transformed using Eq. (4), as was Eq. (13), and the substitution $c_0(t)\approx\overline{c}_0$ then made. But it is more correct to chose $\psi(t,x)$ so as to eliminate $c_0(t)$ or $Dc_x(t,0)$ $(\psi_x(t,0)=0$, $\psi(t,0)=0)$, since the information deficit occurs at the inlet.

When $\psi=\beta(t)\cos\omega x$, $\psi=\beta(t)\sin\omega x$, or $\psi=\zeta(t)\times\exp\omega x$ are used $(\beta(t)=\sigma\exp(D\omega^2t)$, $\zeta(t)=\sigma\exp(-D\omega^2t)$, $\sigma=\text{const})$, the final formulas contain integrals of $\beta(t)J(t)$ and $\beta(t)J^{1/2}(t)$ $(\zeta(t)J(t)$, $\zeta(t)J^{1/2}(t))$. Thus, there is a possibility of giving preference to measurements in the second half of $[0,t_*]$ (greater weight to $\beta(t)$) or in the initial stage of the measurements $(\zeta(t))$. The normalization factor σ is determined, for example, from $\beta(t_*/2)=1$, $\zeta(t_*/2)=1$. The choice of the parameter ω is also important. $D\omega^2$ should not be very large, otherwise a portion of the measurements will be essentially immaterial. The conditions (9) and (10), together with $\psi=x/l$, are satisfied by $\psi=\beta(t)\sin\omega x$. For $\omega=n\pi/l$, both $Dc_x(t,0)$ and $Dc_x(t,l)$ are eliminated from Eq. (8). The equation couples only the concentrations. For $\omega=n\pi/(2l)$ (n odd), the pair $Dc_x(t,0)$, $c_l(t)$ is eliminated. For $\gamma(x)=\cos\omega x$, $\omega=n\pi/l$ we obtain only a relation between the fluxes, while for $\omega=n\pi/(2l)$ (n odd) we obtain a relation between $c_l(t)$ and $Dc_x(t,0)$. If the equation contains $Dc_x(t,0)$ and that quantity is not measured, then we proceed as with Eq. (13). Information on $Dc_x(t,0)$ (or $q_0(t)$, $c_0(t)$), however, greatly increases the reliability of the determination.

We conclude by considering the following possible modification of the experiment. The pressure $\overline{p}_0$ is established by an atomizer (incandescent tungsten filament). We eliminate Eq. (4) from the model, since it presupposes the injection of molecular hydrogen into the vessel. The condition $c_0(t)\approx\overline{c}_0$ ($t\geqslant\varepsilon$, $\varepsilon\ll t_*$) is realized even faster. In the calculations, as before, we replace the unknown value of $\overline{c}_0$ by Eq. (12) ($x=0$). Of course, the levels of $\overline{q}_0$ corresponding to $\overline{p}_0$ should be lower than the theoretical maxima. We do not use Eqs. (15)–(17). Equations of the form (18) and (19), which still contain three unknowns x_1, x_3, and X, remain for determining D, g, and b. In order to remain in the class of linear algebraic equations, it is better to treat x_1, x_3, x_1X, and X as the unknowns. Additional equations can be obtained either by a regular increase in $\overline{p}_0$ ($t=t_*+\Delta t_* \to p=\overline{p}_{03}>\overline{p}_{02}, \overline{J}\to\overline{J}_3, \ldots$), or by repeating steps 1–4 with other $\overline{p}_{01}$ and $\overline{p}_{02}$. Analytical problems do not arise: we write $x_1X=x_2x_3$ from one equation and then solve three linear equations for x_1, x_3, and X.

Actually implementing this technique appears to require a substantial amount of experimental effort. The justification is the (at least, mathematical) nontriviality of the model (1)–(6), which couples the surface and volume processes. Physically, the model is crude; it can be refined. But then the problem of multiparameter determination will be difficult to grasp. The situation becomes simpler if there is already a sufficient number of permeation curves over $[0,t_*]$ for $\varphi(x)=0$ and various $\overline{p}_{0i}$. Then it is possible to get by with just equations of the form $f_2=0$.

COMMENTS

1. If the vacuum system is not powerful enough, the measurements can be modeled by

$$p_l(t)=\theta_1\int_0^t \exp((\tau-t)/\theta_0)J(\tau)d\tau.$$

The flux $J(t)$ is determined uniquely from the pressure $p_l(t)$. A term $\mu s(T)p_l(t)$ is added to the right-hand side of Eq. (5). This leads to no significant changes; the final formulas will contain integrals of $p_l(t)$ as well as of $J(t)$ and $J^{1/2}(t)$.

2. $\overline{J}$ is determined reliably by sufficiently long observation of $J(t)$. But then, a not too large allowed value of t_* should be used in Eq. (11). Otherwise, the informative transient process ($0\to\overline{J}_1$, $\overline{J}_1\to\overline{J}_2$) will be insignificant in the integrals.

3. For large experimental errors it is better to solve numerically the scalar equation (16) rather than Eq. (17) (in particular, by the method of least squares in the real range $s\in[s^-,s^+]$). This can also apply to the linear system of Eqs. (18) and (19), which may be poorly conditioned. For example, with comparatively small $D(T)$ and large $b(T)$ the terms with $x_3=b^{-1/2}$, X in Eq. (18) are relatively small, and Eq. (18) degenerates to $x_1\overline{J}_1/6-\overline{J}_1t_*\sim 0$, i.e., $l^2/(6D)\sim t_*$. The value of x_1 ($D=l^2/x_1$) is determined reliably according to Eq. (20). But the result of trying to determine x_3 is unpredictable if the term $x_3\overline{J}_1^{1/2}$ in Eq. (18) is comparable to the error in determining S_1 and $S_{1/2}$. Besides, the information contained in s, D, and $gb^{-1/2}=lX/D$ is also of practical significance. As $x_3\to 0$ one has $c(t_*,x)\approx 0$ in view of Eq. (12), and the desorption permeation flux degenerates into the diffusion flux; then it is more reasonable to turn to another model. Thus, the problem will be not be well-posed mathematically for all metals: the terms with x_1 and x_3 in Eqs. (18) and (19) must be comparable in order of magnitude (the processes included in the model of Eqs. (1)–(6) must be ''equally important'').

4. In the measurement units taken here the parameters have a large spread in orders of magnitude. Thus, it is appropriate to multiply the equation $f_i=0$ by a scaling factor, say 10^{-12}, and use the new variables $\tilde{J}=J\cdot 10^{-12}$, $\tilde{X}=X\times 10^{-6}$, $\tilde{x}_3=x_3\cdot 10^{-6}$, and $\tilde{x}_4=\mu s\cdot 10^{-12}$.

Thus the method presented here can be used to reduce the nonlinear inverse problem of determining the model parameters of Eqs. (1)–(6) to an analysis of algebraic equations.

This work was supported by the Russian Fund for Fundamental Research (No. 95-01-00355).

[1] G. Alefeld and J. Völkl (Eds.), *Hydrogen in Metals* (Springer, New York, 1978; Mir, Moscow, 1981, Vol. 1, 506 pp., Vol. 2, 430 pp.).

[2] P. V. Gel'd and L. P. Mokhracheva, *Hydrogen and the Physical Properties of Metals and Alloys* [in Russian], Nauka, Moscow (1985), 231 pp.

[3] A. P. Zakharov (Ed.), *Interactions of Hydrogen with Metals* [in Russian], Nauka, Moscow (1987), 296 pp.

[4] I. E. Gabis, A. A. Kurdyumov, and A. V. Samsonov, Pis'ma Zh. Tekh. Fiz. **21**(5), 1 (1995) [Tech. Phys. Lett. **21**, 165 (1995)].

[5] I. E. Gabis, Pis'ma Zh. Tekh. Fiz. **21** (9), 60 (1995) [Tech. Phys. Lett. **21**, 343 (1995)].

[6] O. I. Kon'kov, I. N. Kapitonov, I. N. Trapeznikov, and E. I. Terukov, Pis'ma Zh. Tekh. Fiz. **23**(1), 3 (1997) [Tech. Phys. Lett. **23**, 9 (1997)].

[7] Yu. V. Zaika and I. E. Gabis, Zavod. Lab., No. 1, pp. 18–26 (1996).

[8] I. E. Gabis, A. A. Kurdyumov, and N. A. Tikhonov, Vestn. St. Pb. Univ. Ser. 4, No. 2, p. 77–79 (1993).

[9] Yu. V. Zaika, Zh. Vychil. Mat. Mat. Fiz., No. 12, pp. 108–120 (1996).

[10] G. I. Marchuk, *Conjugate Equations and the Analysis of Complex Systems* [in Russian], Nauka, Moscow (1992), 336 pp.

Translated by D. H. McNeill

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Use of MHD systems in hypersonic aircraft

V. L. Fraĭshtadt, A. L. Kuranov, and E. G. Sheĭkin

O.A.O. Scientific-Research Enterprise for Hypersonic Systems, 196066 St. Petersburg, Russia

(Submitted July 28, 1997)
Zh. Tekh. Fiz. **68**, 43–47 (November 1998)

The possibilities of using magnetohydrodynamic (MHD) systems on hypersonic aircraft are discussed. The distinctive features of using MHD systems in the flow path of ramjet engines are examined. A quasi-one-dimensional mathematical model for the engine is presented which includes the MHD interaction with the flow. It is shown that the specific impulse of an engine system can be raised by using MHD systems. © *1998 American Institute of Physics.* [S1063-7842(98)00811-3]

Hypersonic flight in the atmosphere involves extreme force and thermal effects on the structure of an aircraft. Under certain flight conditions, a plasma ''shell'' can develop around an aircraft, leading to interactions of the aircraft with the surrounding medium which are fundamentally new compared to conventional aerodynamics. Under these conditions, a magneto-gasdynamic volume interaction with the high speed, ionizing flux can be effective for creating control torques, reducing thermal fluxes to the surface of the aircraft, and controlling the structure of the flow.[1] In this paper we examine some distinctive features of using MHD systems in the flow path of a scramjet engine[2] with a magneto-plasma-chemical (MPC) engine developed in the framework of the AJAX concept[3] as an example. The traditional scheme for a scramjet engine has a number of fundamental disadvantages which substantially limit its range of applicability. The complex flow structure in the flow path of a scramjet engine increases the probability of flow separation, which leads to blocking of the channel and makes it more difficult to ignite the fuel in the combustion chamber efficiently.[4] At flight speeds below the design speed, the air intake of a scramjet engine typically has a lower air feed efficiency and a lower degree of compression of the stream. When the speed of a hypersonic aircraft changes, there is a significant realignment of the flow structure in the flow path of the scramjet engine. Altogether, these problems mean that scramjet engines are efficient only within a small range of flight speeds.

In order to extend the domain of operation of scramjet engines, it is necessary to introduce an additional mechanism for acting on the stream which makes it possible to further compress the stream in the air intake, regulate the flow structure, and inhibit the development of separated flows. One of the most promising ways of acting additionally on super- and hypersonic flows in the flow path of ramjets is through a volume interaction using MHD systems. Figure 1 shows a simplified diagram of an MPC engine which implements these principles; it is essentially a scramjet engine with MHD systems inserted in its flow path. Let us examine briefly the functional purpose of the main subsystems of the MPC engine which distinguish it from a scramjet engine. An external MHD generator is used to control the flow profile, regulate the air feed rate in the flow path of the MPC engine, and increase the pressure. An internal MHD generator is used to raise the pressure and prevent the development of separated flows. An ionizer is used to create the required conductivity in the flow when the natural conductivity of the flow does not provide the required degree of MHD interaction. The electrical energy generated by the MHD generators is used to power the ionizer and on-board equipment and to provide further acceleration of the combustion products in the MHD accelerator.

Let us analyze an MPC engine scheme with an internal MHD generator and an MHD accelerator. For clarity we shall do this study with the simplest of assumptions. A quasi-one-dimensional approximation is used in a model of an inviscid, thermally nonconducting ideal gas with a constant specific heat. The MHD flows are described using an approach developed[5,6] for analyzing complex systems, including MHD systems. Let us examine the features of this approach briefly. Formally assuming that the pressure gradient in the MHD channel is proportional to the force exerted on the flow by the magnetic field, we introduce a proportionality coefficient ξ. For an ideally sectored Faraday MHD channel we assume that

$$\frac{dp}{dx}=\xi(x)(1-k)^2\sigma B^2 v, \tag{1}$$

where p is the static pressure in the flow, v is the flow velocity, x is the longitudinal spatial coordinate, k is the load coefficient, σ is the conductivity of the flow, and B is the magnetic induction.

If we limit ourselves to the class of solutions for which ξ is constant, then, using Eq. (1), we can obtain simple analytical expressions for the parameters at the outlet of the MHD channel. The corresponding flow regime will be called the $\xi=\mathrm{const}$ flow regime. The changes in the flow parameters in the MHD channel are given by

$$\frac{T_2}{T_1}=1+\frac{1-k}{k}\left(1+\frac{\gamma-1}{2}M_1^2\right)(1+\xi)\,\eta,$$

$$\frac{v_2}{v_1}=\sqrt{1-\frac{1+\xi(1-k)}{k}G\,\eta},$$

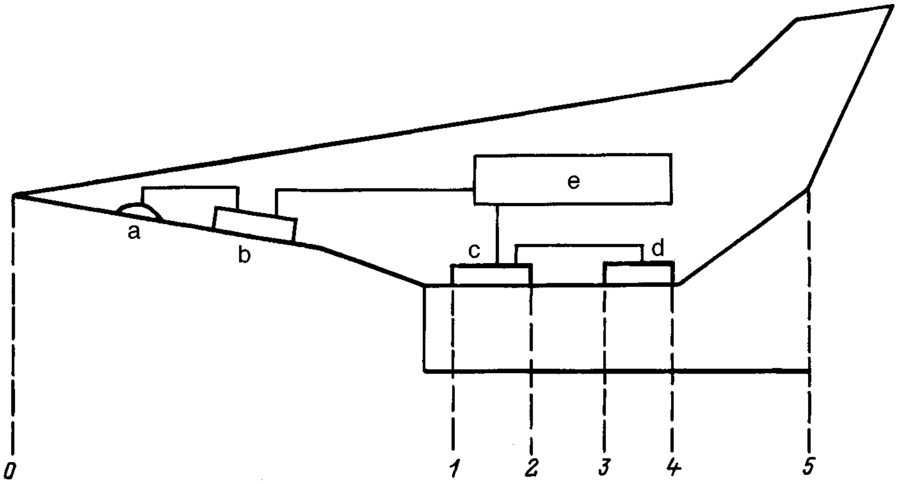

FIG. 1. Simplified sketch of a magnetoplasma-chemical engine: (*0–1*) air intake, (*1–2*,c) internal MHD generator, (*2–3*) combustion chamber, (*3–4*, d) MHD accelerator, (a) ionizer, (b) external MHD generator, (e) onboard systems.

$$\frac{p_2}{p_1}=\left(\frac{T_2}{T_1}\right)^{\frac{\gamma}{\gamma-1}\frac{\xi}{\xi+1}},\quad G=\frac{2+(\gamma-1)M_1^2}{(\gamma-1)M_1^2}. \tag{2}$$

Here T is the temperature, M is the Mach number, γ is the adiabatic index, η is the coefficient of conversion of the enthalpy of the flow into electrical energy, and the subscripts 1 and 2 label the parameters at the inlet and outlet of the MHD channel, respectively. The $\xi=\text{const}$ flow regime includes, as a special case, the often examined flow regimes that are characterized by conservation of one of the flow parameters. The values of ξ corresponding to these regimes are listed in Table I.

In analyzing an MPC engine with an internal MHD generator, we shall use the following subscripts to denote the parameters at various locations: 0 in the incident flow, 1 at the entrance to the MHD generator, 2 at the entrance to the combustion chamber, 3 at the entrance to the MHD accelerator, 4 at the entrance to the jet nozzle, and 5 at the outlet of the nozzle. (Naturally, the outlet parameters of a subsystem are the inlet parameters of the subsystem located after it.)

In this paper we limit ourselves to examining the case in which conductivity of the flow is achieved without the use of an ionizer. We examine the subsystems of an MPC engine and determine the relationships among the parameters at the inlet and outlet of the system.

The air intake includes an external part, which compresses the entering flow in a system of oblique shocks, and an internal part (isolator), which provides for a return and further compression of the flow. The following characteristics are used: N, the number of shocks in the external part, Θ_N, the net return flux in the air intake, and σ_{in}, the coefficient of restitution of the total pressure in the air intake. If the temperature at the outlet of the air intake (the inlet of the MHD generator) is T_1, then the changes in the pressure and velocity in this subsystem are determined by the following equations:

$$\frac{p_1}{p_0}=\sigma_{\text{in}}\left(\frac{T_1}{T_0}\right)^{\frac{\gamma}{\gamma-1}},\quad \frac{v_0^2}{2}+c_pT_0=\frac{v_1^2}{2}+c_pT_1, \tag{3}$$

where c_p is the specific heat of air.

The MHD generator is characterized by the parameters ξ_1 and k_1 and the enthalpy conversion coefficient η. The changes in the flow parameters in the channel of the MHD generator are determined by

$$\frac{T_2}{T_1}=1+\frac{1-k_1}{k_1}\left(1+\frac{\gamma-1}{2}M_1^2\right)(1+\xi_1)\eta,$$

$$\frac{p_2}{p_1}=\left(\frac{T_2}{T_1}\right)^{\frac{\gamma}{\gamma-1}\frac{\xi_1}{\xi_1+1}}. \tag{4}$$

We consider a combustion chamber operating at constant pressure. Since the mass feed rate of fuel is usually much lower than that of air, we shall treat the delivery of fuel to the combustion chamber as heat release without mass input. Then the changes in the flow parameters in the combustion chamber have the simpler form^2

$$T_3=T_2+\Delta T,\quad p_3=p_2,$$

$$\Delta T=\frac{H_u}{c_p(\alpha L_0+1)}, \tag{5}$$

where H_u is the calorific value of the fuel, L_0 is the stoichiometric coefficient, and α is the excess air factor.

The MHD accelerator is characterized by the parameters ξ_3 and k_3. It is assumed that all the energy produced by the MHD generator is transferred to the MHD accelerator. The changes in the flow parameters in the MHD accelerator channel are determined by

TABLE I.

Flow regime	Value of ξ corresponding to the given flow regime
$\rho=\text{const}$	$\xi=\gamma-1$
$p=\text{const}$	$\xi=0$
$T=\text{const}$	$\xi=-1$
$M=\text{const}$	$\xi=-\left[1+\frac{2}{(1-k)(\gamma-1)M_1^2}\right]\frac{1}{G}$
$v=\text{const}$	$\xi=-1/(1-k)$

$$\frac{T_4}{T_3}=1+\frac{k_3-1}{k_3}\left(1+\frac{\gamma-1}{2}M_1^2\right)\frac{T_1}{T_3}(1+\xi_3)\,\eta,$$

$$\frac{p_4}{p_3}=\left(\frac{T_4}{T_3}\right)^{\frac{\gamma}{\gamma-1}\,\frac{\xi_3}{\xi_3+1}}. \tag{6}$$

We assume that the flow in the nozzle is isentropic. Then the relative change in the flow pressure is related to the relative temperature change by

$$\frac{p_5}{p_4}=\left(\frac{T_5}{T_4}\right)^{\frac{\gamma}{\gamma-1}}. \tag{7}$$

At the design efflux from the nozzle, the system of Eqs. (3)–(7) can be closed by assuming that the pressure at the nozzle exit is the same as in the surrounding medium, i.e., $p_5=p_0$. Given this relationship, the system of Eqs. (3)–(7) yields the following formula for calculating the flow temperature at the nozzle exit:

$$T_5=\frac{T_4}{\sigma_{\text{in}}^{(1-1/\gamma)}\left[\frac{T_1}{T_0}\left(\frac{T_2}{T_1}\right)^{\frac{\xi_1}{\xi_1+1}}\left(\frac{T_4}{T_3}\right)^{\frac{\xi_3}{\xi_3+1}}\right]}. \tag{8}$$

The efflux velocity of the gas from the nozzle is determined in terms of the temperature T_5 using the conservation of energy,

$$v_5=\sqrt{v_0^2+2c_p(T_0+\Delta T-T_5)}. \tag{9}$$

These formulas can be used to determine the specific impulse I_{sp} of the MPC engine. Neglecting the mass feed rate of fuel compared to that of air, we obtain[7]

$$I_{\text{sp}}=\frac{\alpha L_0}{g}(\varphi v_5-v_0), \tag{10}$$

where g is the acceleration of gravity and φ is a coefficient which takes the nonideality of the nozzle into account.

In those cases where it is not specially noted otherwise, we shall set $\varphi=1$.

The set of Eqs. (3)–(10) can be used to calculate the specific impulse of the MPC engine for given parameters of the air intake, MHD system, and combustion chamber. Here the specific impulse depends on a large number of parameters: α, L_0, M_0, T_1, σ_{in}, k_1, ξ_1, η, k_3, and ξ_3. T_1 and σ_{in} are defined in terms of the air intake parameters N and Θ_N and a computational technique similar to that described in Ref. 8 was used, with posterior averaging of the parameters in the outlet section of the air intake. We have determined the range of variation of the parameters of the subsystems of an MPC engine within which the use of an MHD system makes it possible to increase the specific impulse of the engine system. We use the obvious functional relationship

$$\left.\frac{\partial I_{\text{sp}}}{\partial\eta}\right|_{\eta\to 0}>0.$$

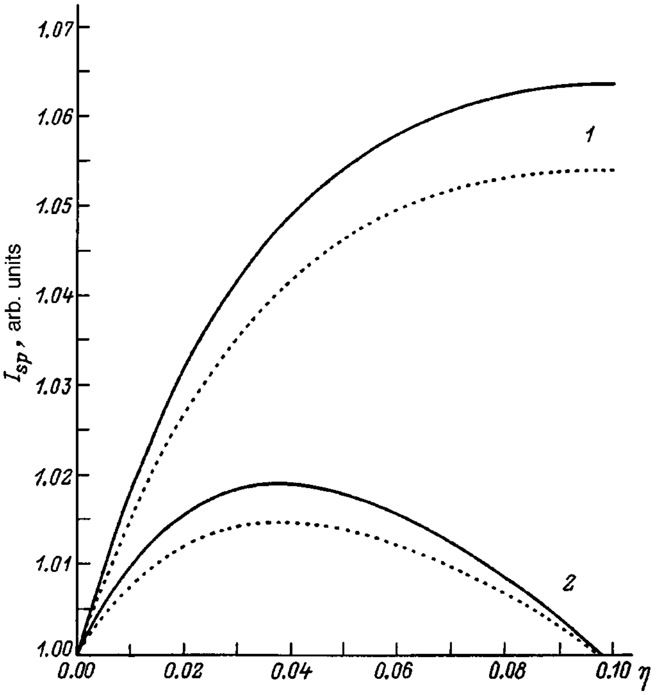

FIG. 2. Specific impulse of a magneto-plasma-chemical engine as a function of the conversion efficiency of flow enthalpy into electrical energy: $\Theta_N=0.2$ rad, $k_1=0.5$, $k_3=2$; $M_0=6$ (*1*), 8 (*2*); smooth curves $\varphi=0.95$, dotted curve $\varphi=1$.

Equations (9) and (10) imply that this condition is equivalent to the condition

$$\left.\frac{\partial T_5}{\partial\eta}\right|_{\eta\to 0}<0.$$

After the required transformations, we obtain the following inequality:

$$\xi_1>\frac{T_1}{\Delta T}\,\frac{1-k_1/k_3}{1-k_1}. \tag{11}$$

Since the load coefficient for the MHD generator is $0<k_1<1$ and for the MHD accelerator $k_3>1$, the specific impulse of an MPC engine in this configuration increases for positive ξ_1, which, according to Eq. (1), corresponds to an MHD generator operating with an elevated pressure along the channel length. The requirements on the magnitude of the pressure drop are less at higher ΔT and lower T_1. Figure 2 shows the specific impulse of the MPC engine as a function of the coefficient of conversion of the enthalpy of the flow to electrical energy for different values of the Mach number of the incident flow for ideal and nonideal nozzles. A value $\eta=0$ corresponds to a scramjet engine. All these curves are normalized to the specific impulse of a scramjet engine. (The curves of Figs. 2–5 are for $\xi_1=\xi_3=\alpha=1$ and $N=2$.)

In all the calculated variants, MHD energy conversion in the flow path of the engine system leads to an increase in the specific impulse, and for a nonideal nozzle the positive effect is more significant. The relative increase in the specific impulse of the MPC engine in this variant is more significant for lower Mach numbers.

The dependence of the specific impulse on the load coefficient of the MHD generator shown in Fig. 3 is nonmono-

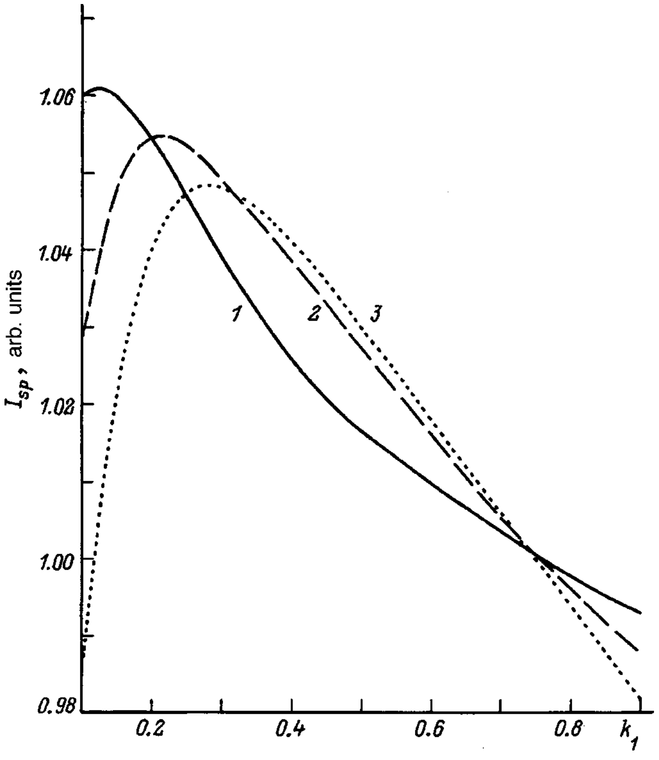

FIG. 3. Specific impulse of a magneto-plasma-chemical engine as a function of the load coefficient of the MHD generator: $M_0=6$, $\Theta_N=0.2$ rad, $k_3=2$; $\eta=0.05$ (*1*), 0.1 (*2*), 0.15 (*3*).

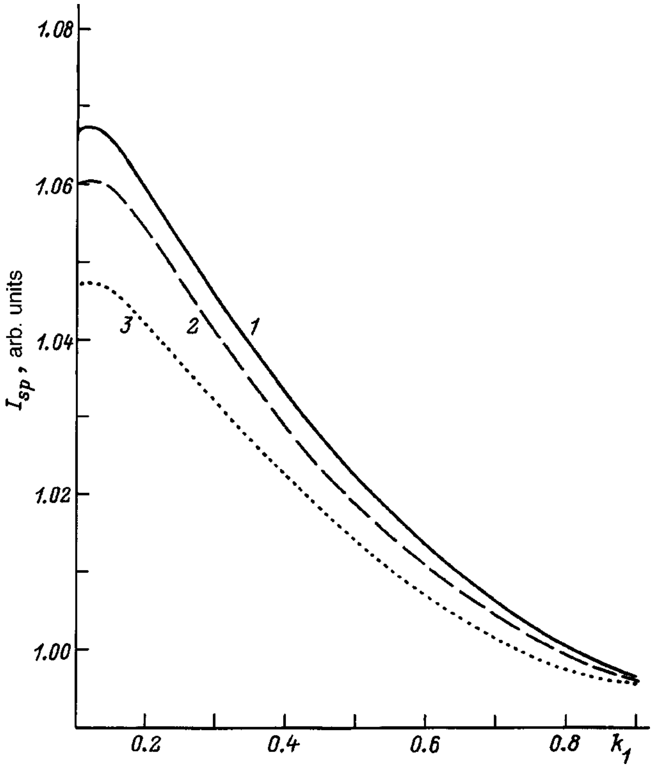

FIG. 5. As in Fig 3, for $M_0=6$, $k_3=2$, $\eta=0.05$; $\Theta_N=0.1$ (*1*), 0.2 (*2*), 0.3 rad (*3*).

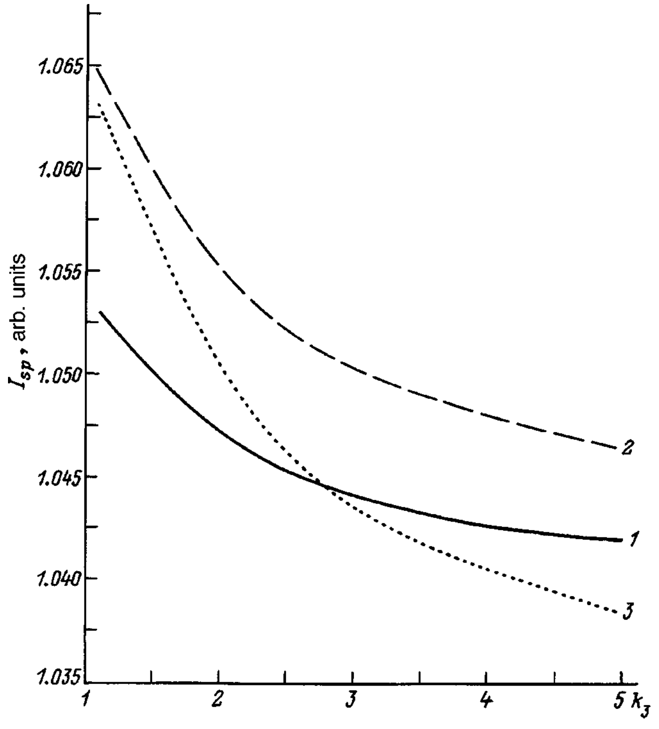

FIG. 4. Specific impulse of a magneto-plasma-chemical engine as a function of the coefficient k_3: $M_0=6$, $\Theta_N=0.2$ rad, $k_1=0.25$, $\alpha=1$, $\eta=0.05$ (*1*), 0.1 (*2*), 0.15 (*3*).

tonic, with a distinct extremum. The magnitude and location of the extremum depend on the conversion coefficient of flow enthalpy into electrical energy. With increasing η the extremum shifts toward larger k_1, while its magnitude decreases. Figure 4 shows that the specific impulse of an MPC engine falls off monotonically with rising k_3. With increasing η the dependence of the specific impulse on the load coefficient k_3 becomes more pronounced. The results shown in Fig. 5 imply that the relative increase in the specific impulse is maximum for an MPC engine with an air intake characterized by a minimum turn angle for the flow.

These calculations show that using MHD systems in the flow path of a scramjet engine with a suitable choice of parameters makes it possible to increase the specific impulse of the engine system. We have found the limits on the range of variation in the parameters of the MPC engine subsystems that will ensure enhanced specific impulse for the system. In later papers we shall examine the possibility of using MHD interactions for controlling the flow structure and study the characteristics of MPC engines in a two-dimensional Euler approximation.

[1] V. A. Bityurin, J. T. Linebery *et al.*, ''Assessment of hypersonic MHD concepts,'' AIAA paper 97-2393.

[2] R. I. Kurziner, *Jet Engines for High Supersonic Flight Speeds* [in Russian], Mashinostroenie, Moscow (1989), 264 pp.

[3] E. P. Gurijanov and P. T. Harsha, AJAX: New Directions in Hypersonic Technology, AIAA paper 96-4609.

[4] E. H. Andrews, C. A. Trexler, and S. Emami, Tests of a Fixed-Geometry Inlet-Combustor Configuration for a Hydrocarbon Fueled Dual-Mode Scramjet, AIAA 94-2817.

[5] E. G. Sheĭkin, Zh. Tekh. Fiz. **62**(12), 1 (1992) [Sov. Phys. Tech. Phys. **37**, 1133 (1992)].

[6] E. G. Sheĭkin, Zh. Tekh. Fiz. **63**(9), 6 (1993) [Tech. Phys. **38**, 741 (1993)].

[7] E. T. Curran, J. Leingang, L. Carreiro, and D. Peters, Further Studies of Kinetic Energy Methods in High Speed Ramjet Cycle Analysis, AIAA 92-3805.

[8] A. G. Berlyand and V. I. Penzin, TsAGI Preprint No. 11 [in Russian], Central Aerohydrodynamic Institute, Moscow (1990), 20 pp.

Translated by D. H. McNeill

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Stability of a charged drop having the form of a triaxial ellipsoid

S. I. Shchukin and A. I. Grigor'ev

Yaroslavl State University, 150000 Yaroslavl, Russia

(Submitted November 3, 1997)
Zh. Tekh. Fiz. **68**, 48–51 (November 1998)

The stability of a highly charged, isolated conductive drop is analyzed within the principle of minimum potential energy of a closed system. A treatment of the stability of drops of ellipsoidal shape shows that both spherical drops and drops having an oblate spheroidal shape experience instability at sufficiently large charges according to a single scheme, i.e., they deform to a prolate spheroid. © *1998 American Institute of Physics.* [S1063-7842(98)00911-8]

The investigation of the stability of charged drops is of interest for many areas of physics and technology: from the liquid drop model of the nucleus and explanations for the shape of planets to the development of technologies for the electrospraying of liquids and determination of the heat- and mass-transfer coefficients in heterogeneous media.[1–3] The history of the research pertaining to this subject dates back to the end of the last century, when Rayleigh published a paper[4] in which he showed that a spherical drop becomes unstable at a certain relationship between surface-tension and electrostatic forces. It was also shown in Ref. 4 that for small perturbations of the shape of a spherical drop, the fundamental axisymmetric mode, which is proportional to the second Legendre polynomial, i.e., $\propto P_2(\cos\Theta)$, has the lowest excitation energy. For this reason, only axisymmetric shapes have been considered in more recent studies devoted to the stability of charged drops.[1] In particular, the investigations of different authors employing diverse approaches to finding the stable shapes of a charged drop[5–7] have led to the conclusion that a highly charged drop in the form of an axisymmetric oblate spheroid of revolution is stable, a conclusion which is not entirely clear from general physical arguments.

The purpose of the present work is to investigate the stability of charged ellipsoidal drops and the laws governing the onset of their instability against a self-charge. We write the equation of the free surface of an ellipsoidal drop in the form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1,$$

where $a>b>c$ are the semiaxes.

In the limit $a=b>c$ we have an oblate axisymmetric ellipsoid, and for $a>b=c$ we have a prolate axisymmetric ellipsoid.

The total potential energy of a conductive charged ellipsoidal drop is determined by the sum of the energy of the forces of surface tension and the electrostatic energy of the charge Q on the drop,[5,6]

$$U_E=S\sigma+\frac{Q^2}{2C},$$

where C is the capacitance and σ is the surface tension.

We write the surface area S of an ellipsoid in the form

$$S=2\pi c^2+\frac{2\pi b}{\sqrt{a^2-c^2}}\{c^2\mathbf{F}(m,k)+(a^2-c^2)\mathbf{E}(m,k)\},$$

where

$$m=\frac{\sqrt{a^2-c^2}}{a},\quad k=\frac{a}{b}\sqrt{\frac{b^2-c^2}{a^2-c^2}},$$

$\mathbf{F}$ is an elliptic integral of the first kind, and $\mathbf{E}$ is an elliptic integral of the second kind.[8]

The capacitance of an ellipsoid is given by the known relation[9]

$$\frac{1}{C}=\int_0^\infty\frac{1}{\sqrt{(a^2+x)(b^2+x)(c^2+x)}}\,dx,$$

and it can also be expressed in terms of elliptic integrals:[9]

$$\frac{1}{C}=\frac{\mathbf{K}\left(\sqrt{\dfrac{a^2-b^2}{a^2-c^2}}\right)-\mathbf{F}\left(\dfrac{c}{b},\sqrt{\dfrac{a^2-b^2}{a^2-c^2}}\right)}{\sqrt{a^2-c^2}}.$$

If we introduce two parameters (x and y), which uniquely characterize the shape of an ellipsoid,

$$x=\frac{a}{b},\quad y=\frac{b}{c},$$

and write the equation relating the values of the semiaxes of the ellipsoids to the radius of an equivalent sphere $R^3=abc$, then expressing a, b, and c in terms of x, y, and R

$$a=R(x^2y)^{\frac{1}{3}},\quad b=R\left(\frac{y}{x}\right)^{\frac{1}{3}},\quad c=\frac{R}{(xy^2)^{\frac{1}{3}}},$$

we can write the total potential energy of a charged ellipsoidal drop in the form

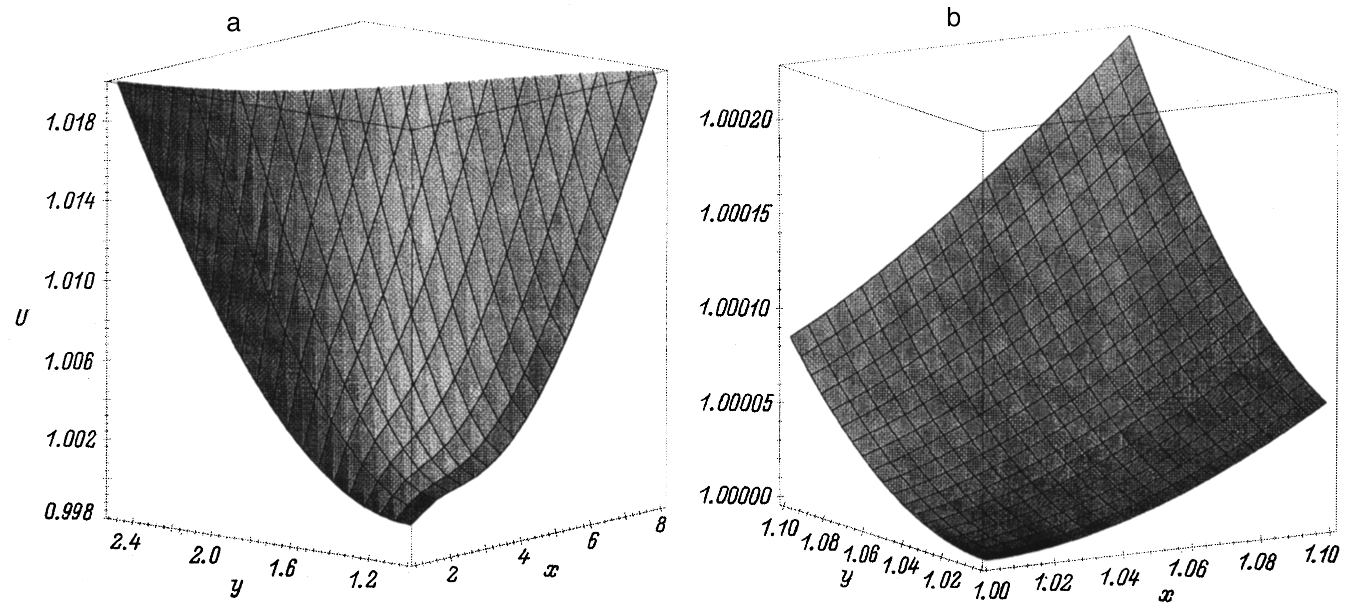

FIG. 1. a — Dependence of the dimensionless energy of a charged ellipsoidal drop on the ratio between the semiaxes x and y, which characterize the geometry of a spheroid, for $W=3.5$; b — same dependence, but on a magnified scale in the vicinity of the point $x=1$, $y=1$, i.e., for very small ellipsoidal deformations of the sphere.

$$U_E=2\pi\sigma R^2\left\{\sqrt{x^2y^2-1}+y[\mathbf{F}(m,k)+(x^2y^2-1)\mathbf{E}(m,k)]+Wxy^2\times\left[\mathbf{K}(p)-\mathbf{F}\left(\frac{1}{y},p\right)\right]\right\}\cdot[x^{2/3}y^{4/3}\sqrt{x^2y^2-1}]^{-1};$$

$$W=\frac{Q^2}{4\pi R^3\sigma};\quad m=\sqrt{1-\frac{1}{x^2y^2}};$$

$$k=\frac{x\sqrt{y^2-1}}{\sqrt{x^2y^2-1}};\quad p=\frac{y\sqrt{x^2-1}}{\sqrt{x^2y^2-1}}.$$

We render this expression dimensionless by dividing it by the potential energy of a charged conductive spherical drop of equivalent volume

$$U_S=4\pi R^2\sigma\left(1+\frac{1}{2}W\right)$$

and obtain

$$U=\left\{\sqrt{x^2y^2-1}+y\left[\mathbf{F}\left(\sqrt{1-\frac{1}{x^2y^2}},\frac{x\sqrt{y^2-1}}{\sqrt{x^2y^2-1}}\right)+(x^2y^2-1)\mathbf{E}\left(\sqrt{1-\frac{1}{x^2y^2}},\frac{x\sqrt{y^2-1}}{\sqrt{x^2y^2-1}}\right)\right]+Wxy^2\left[\mathbf{K}\left(\frac{y\sqrt{x^2-1}}{\sqrt{x^2y^2-1}}\right)-\mathbf{F}\left(\frac{1}{y},\frac{y\sqrt{x^2-1}}{\sqrt{x^2y^2-1}}\right)\right]\right\}\times[x^{2/3}y^{4/3}\sqrt{x^2y^2-1}(W+2)]^{-1};\quad U=\frac{U_E}{U_S}.$$

Plots of the dependence of the dimensionless potential energy of an ellipsoidal drop $U=U(x,y)$ on the ratio between the semiaxes for various values of the drop charge (the Rayleigh parameter W) are shown in Figs. 1, 2, 3, and 4. The $x=1$ plane corresponds to an oblate ellipsoid of revolution, and the $y=1$ plane corresponds to a prolate ellipsoid of revolution. The point $x=1$, $y=1$ corresponds to a spherical drop, whose energy, as follows from the normalization condition, equals unity.

The plot of $U=U(x,y)$ for $W=3.5$ (Figs. 1a and 1b) shows that the energy of an ellipsoidal drop is greater than the energy of a spherical drop when $W=3.5$ and that the energy of the drop increases with increasing degree of deformation. Such a tendency is manifested both at large deformations (Fig. 1a) and at small deformations (Fig. 1b). Thus, when $W\leqslant 3.5$, the spherical shape is stable.

When $W=3.9$ (Figs. 2a and 2b), the dependence of $U=U(x,y)$ has a somewhat different character: in the vicinity of the point $x=1$, $y=1$ weak deformation of the drop leads to an increase in its energy (Fig. 2b), while at considerable degrees of deformation the $U=U(x)$ curve displays an energy minimum for a prolate spheroid at $x=4.9$ (Fig. 2a). A more detailed investigation reveals that the minimum on the plot of $U=U(x)$, i.e., the energy of a drop deformed to a prolate spheroid, appears when $W>3.546$.

When $W\geqslant 4$, the plot of $U=U(x,y)$ (Figs. 3a and 3b, which were calculated for $W=4.1$) not only has the minimum at large deformations that was observed for $W=3.9$ (Fig. 2a), but, as expected, has a decreasing course at

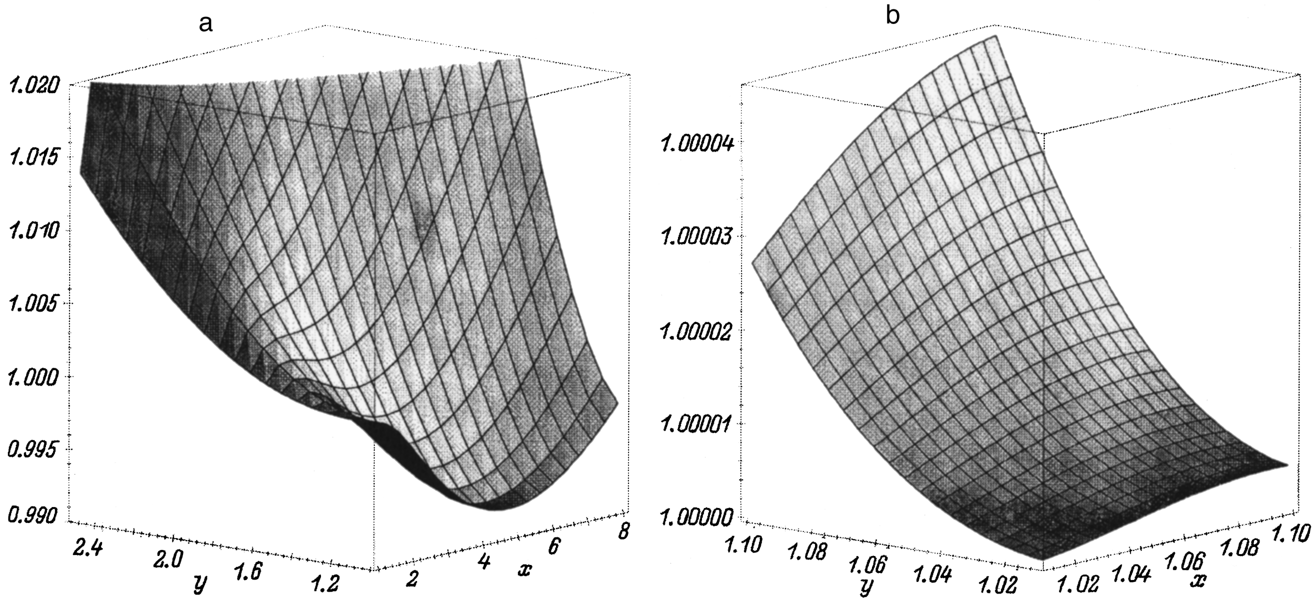

FIG. 2. a — Dependence of the dimensionless energy of a charged ellipsoidal drop on the ratio between the semiaxes *x* and *y*, which characterize the geometry of a spheroid, for $W=3.9$; b — same dependence, but on a magnified scale in the vicinity of the point $x=1$, $y=1$, i.e., for very small ellipsoidal deformations of the sphere.

small deformations, and the drop becomes unstable against infinitesimal virtual changes in energy (Fig. 3b).

A comparison of Figs. 2a and 3a reveals that the position for the minimum on the $U=U(x)$ curve shifts toward increasing values of *x* as *W* is increased.

An investigation of $U=U(x,y)$ in the region of the energy minimum for an oblate spheroid shows that this shape is unstable, as follows from Fig. 4a, which was calculated for $W=4.3$. It is seen that the transition from the oblate spheroid with minimum energy to a triaxial ellipsoid is energetically favorable, because it leads to a further decrease in the energy of the drop. A highly charged drop having the form of a triaxial ellipsoid is also unstable, and its shape evolves into the prolate spheroid having the minimum energy. A prolate spheroidal shape is energetically most favorable for a highly charged drop, as can be seen from Fig. 4b. When $W>4$ the

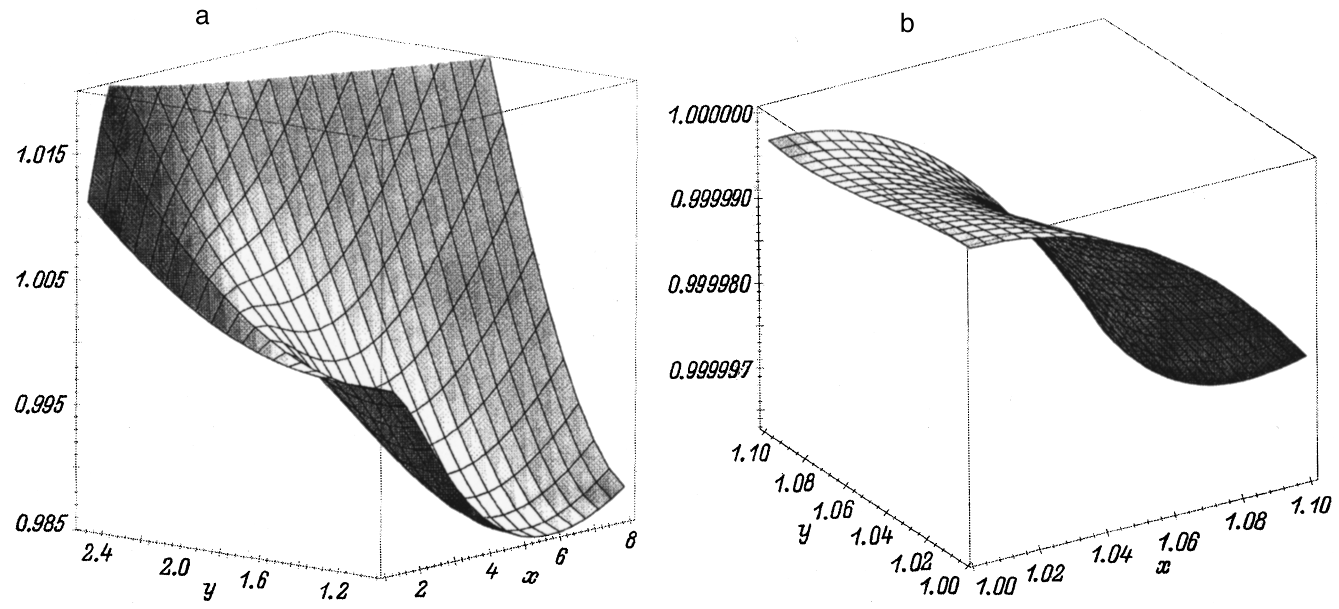

FIG. 3. a — Dependence of the dimensionless energy of a charged ellipsoidal drop on the ratio between the semiaxes *x* and *y*, which characterize the geometry of a spheroid, for $W=4.1$; b — same dependence, but on a magnified scale in the vicinity of the point $x=1$, $y=1$, i.e., for very small ellipsoidal deformations of the sphere.

minimum of the function $U(x,1)$ is the global minimum of the function $U(x,y)$, i.e., deformation of the minimum-energy prolate spheroid along the direction perpendicular to the symmetry axis leads to an increase in the energy of the drop.

A charged spherical drop is stable against infinitesimal perturbations of its shape when $W<4$. In the range of values of the Rayleigh parameter $3.546<W<4$ the plots of $U=U(x)$ display not only the minimum at $x=1$, but also another minimum at $x\geqslant 3$; therefore, sufficiently energetic external disturbances can lead to bifurcation of the drop shape. A spherical drop is unstable at values of the Rayleigh parameter $W>4$. The state of a drop in the form of an oblate spheroid is unstable at any value of the Rayleigh parameter: when $W<4$, the energy of such a drop is greater than that of a spherical drop, and when $W>4$, a drop having the shape of an oblate spheroid stretches along one of the directions perpendicular to the symmetry axis, thereby transforming into a triaxial ellipsoid and then into a prolate spheroid, for which the energy is minimum.

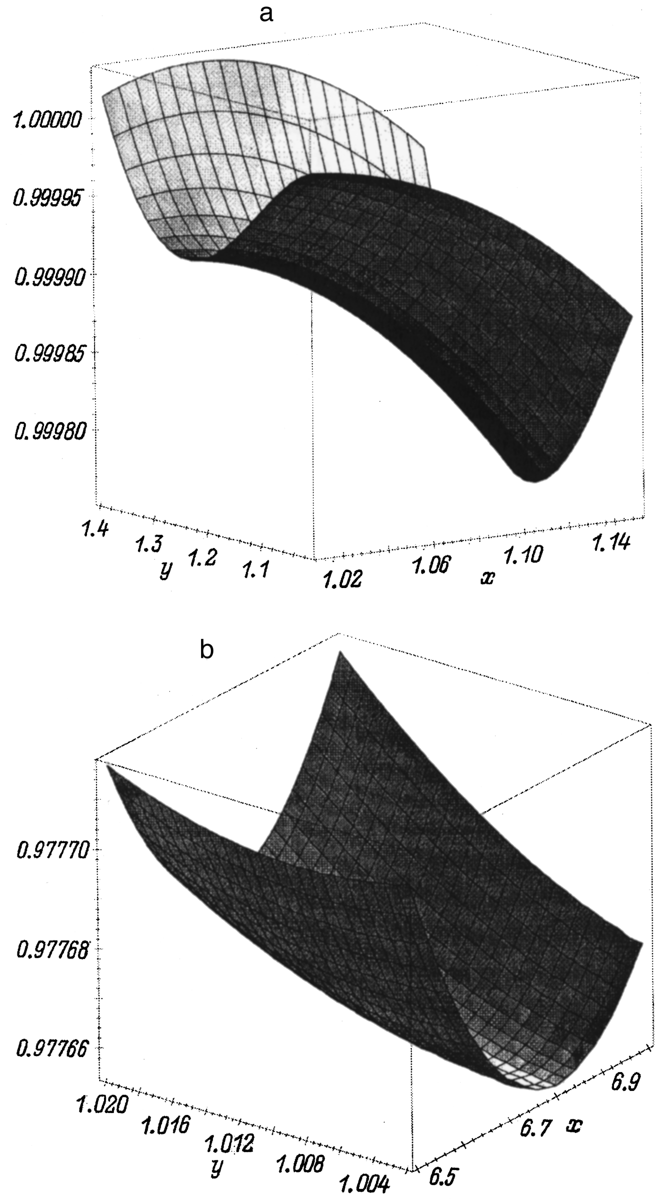

FIG. 4. Variation of the energy of a drop having the shape of an oblate (a) and a prolate (b) spheroid for deformation along a direction perpendicular to the symmetry axis for $W=4.3$.

[1] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).

[2] S. I. Bastrukov, Phys. Rev. E **53**, 1917 (1996).

[3] S. I. Bastrukov and I. V. Molodtsova, Dokl. Ross. Akad. Nauk **350**, 321 (1996) [Phys. Dokl. **41**, 388 (1996)].

[4] Lord Rayleigh (J. W. Strutt), Philos. Mag. **14**, 184 (1882).

[5] G. Ailam and I. Gallily, Phys. Fluids **5**, 575 (1962).

[6] O. A. Bassaran and L. E. Scriven, Phys. Fluids A **1**, 795 (1989).

[7] A. I. Grigor'ev, A. A. Firstov, and S. O. Shiryaeva, in *Proceedings of the 9th International Conference on Atmospheric Electricity*, St. Petersburg (1992), pp. 450–453.

[8] A. Erdélyi (Ed.), *Bateman Manuscript Project. Higher Transcendental Functions, Vol. 3*, (McGraw-Hill, New York, 1955; Nauka, Moscow, 1955, 299 pp.)

[9] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, [Pergamon, Oxford (1984); Mir, Moscow (1982), 620 pp.].

Translated by P. Shelnitz

Stochastic heating in a plasma–beam system

O. V. Klimov and A. A. Tel'nikhin

Altai State University, 656058 Barnaul, Russia

(Submitted March 11, 1997)
Zh. Tekh. Fiz. **68**, 52–56 (November 1998)

The excitation of a Langmuir plasma wave by a monoenergetic electron beam for which resonant interaction conditions hold is investigated within a hydrodynamic description of the plasma. It is shown that parametric and modulation effects lead to the formation of nonlinear stationary waves with a low-frequency soliton-like envelope in the plasma. The behavior of electrons in the field of the wave packet formed by Langmuir waves with different phase velocities is investigated. The level of stochasticity in the system and the relative level of plasma fluctuations are determined. © *1998 American Institute of Physics.* [S1063-7842(98)01011-3]

INTRODUCTION

The results of experiments devised to investigate the evolution of the parameters of a plasma–beam system were described in Refs. 1–3. It was discovered that short flashes of electromagnetic radiation are generated in the plasma at a frequency near the electron plasma frequency Ω_e. It was shown that the radiation sources are localized in space, and their characteristic dimension was determined. The formation of the electron distribution function was recorded experimentally in Ref. 2, where measurements of the emission spectra in the low-frequency region were performed and the fluctuation spectrum was investigated by a probe method.

The experimental data obtained by Karfidov *et al.*[2] were interpreted as being a result of the formation of strong Langmuir turbulence in the plasma–beam system, whose pattern can be described by the phenomenological theory for an ensemble of collapsing Langmuir cavitons.

In this paper it is shown that some basic features of the evolution of a plasma–beam system can be described as being aresult of the generation of nonlinear Langmuir waves during the induced scattering of beam electrons. The attendant beam instability, which is accompanied by self-modulation and bunching of the electron beam, is stabilized by the trapping of beam particles by the waves with the resultant formation of stationary nonlinear waves.[4] The dynamics of electrons in such a wave field become complicated, and dynamical chaos arises in the system for certain values of the parameters.

DYNAMICS OF PLASMA WAVES

Let a fairly intense, nonrelativistic electron beam propagate in an isotropic plasma. In the equilibrium state the ratio $n_b/n_0 \ll 1$, where n_0 is the plasma density and n_b is the beam density, and the beam velocity $V_0 \gg V_T$, where V_T is the thermal velocity of the electrons. Because the charge-density waves existing in the plasma at the thermal level lead to modulation of the beam and thereby intensify the modulating wave, the beam and plasma parameters fluctuate about the equilibrium values. The plasma parameters are denoted by n_j and v_j ($j=1,2$; the subscripts 1 and 2 refer to plasma and beam parameters; n_1/n_0, n_2/n_b, $v_2/V_0 \ll 1$).

The evolution of n_j and v_j for a plasma–beam system will be described by hydrodynamic equations and Poisson's equation for the electric field E:

$$\frac{\partial n_j}{\partial t} + \frac{\partial}{\partial x}(n_j v_j) = 0,$$

$$\frac{\partial v_j}{\partial t} + v_j \frac{\partial v_j}{\partial x} = -\frac{eE}{m_e}, \quad \frac{\partial E}{\partial x} = -4\pi e \sum_j n_j. \tag{1}$$

From (1) we can obtain equations which describe the dynamics of nonlinear density waves:

$$\left(\frac{\partial^2}{\partial t^2} + \Omega_e^2\right) n_1 + \frac{\partial^2 n_1 v_1}{\partial x \partial t} - \frac{n_0}{2}\frac{\partial^2 v_1^2}{\partial x^2} + \Omega_e^2 n_2 = 0,$$

$$\left(\frac{\partial^2}{\partial t^2} + 2V_0 \frac{\partial^2}{\partial x \partial t} + V_0^2 \frac{\partial^2}{\partial x^2} - \Omega_b^2\right) n_2 + V_0 \frac{\partial^2 n_2 v_2}{\partial x^2}$$

$$+ \frac{\partial^2 n_2 v_2}{\partial x \partial t} - \frac{n_b}{2}\frac{\partial^2 v_2^2}{\partial x^2} + \Omega_b^2 n_1 = 0. \tag{2}$$

The system of equations (2) corresponds to the ordinary dispersion relation

$$1 = \frac{\Omega_e^2}{\omega^2} + \frac{\Omega_b^2}{(\omega - kV_0)^2}, \quad \Omega_b^2 = \frac{4\pi e^2 n_b}{m_e}.$$

It follows from an analysis of this equation that waves which are unstable when an electron beam interacts with a plasma satisfy the conditions

$$\omega \leqslant kV_0, \quad \omega = \Omega_e. \tag{3}$$

Assuming that high harmonics are generated during a nonlinear interaction, we substitute the expansions of n_j and

v_j in the form of series into (2). Using the conditions (3) and carrying out the averaging over the high-frequency modes, we find equations which describe the slow evolution of the complex amplitudes a_j $[n_j=a_j\exp(ikx-i\omega t)]$ of charge-density waves:

$$\begin{aligned} i\dot{a}_1&=\alpha_1|a_1|^2a_1+\beta_1|a_2|^2a_1+\lambda_1a_2,\\ -i\dot{a}_2&=\alpha_2|a_2|^2a_2+\beta_2|a_1|^2a_2+\lambda_2a_1. \end{aligned}\tag{4}$$

Here the dot denotes differentiation with respect to the "new" time $t'=t-x/V_0$. Going over to the new variable t' essentially means that from the entire set of solutions for a_j, we are selecting solutions in the form of stationary waves moving with a velocity V_0. Such waves have been detected repeatedly in experiments (see, for example, Refs. 1 and 3). The coefficients α_1 and β_2 reflect the character of the nonlinear interaction, and $(\lambda_1\lambda_2)^{1/2}=\gamma$ is the linear beam instability growth rate [in deriving (4) it was assumed that $\gamma\ll\Omega_e$]. In (4) these coefficients have the following form:

$$\begin{aligned} &\alpha_1=2\frac{\Omega_e}{n_0^2},\quad \beta_1=\frac{3}{4}\frac{\Omega_e}{n_0^2}\left(\frac{\Omega_e}{\Omega_b}\right)^2,\\ &\alpha_2=3\frac{\Omega_e}{n_0^2}\left(\frac{\Omega_e}{\Omega_b}\right)^2,\quad \beta_2=-\frac{3}{8}\frac{\Omega_e}{n_0^2},\\ &\lambda_1=\frac{1}{2}\Omega_e,\quad \lambda_2=\frac{1}{2}\frac{\Omega_b^2}{\Omega_e},\quad \gamma=\frac{\Omega_e}{2}\left(\frac{n_b}{n_0}\right)^{1/2}. \end{aligned}\tag{5}$$

The maximum instability growth rate for a monoenergetic beam $\gamma_{\max}\approx\Omega_e(n_b/n_0)^{1/3}$. In the general case the conditions $(\pi/2)^{1/2}(n_b/n_0)\Omega_e(\Omega_e/kV_{Tb})^2<\gamma<\gamma_{\max}$, where V_{Tb} is the thermal spread in the beam, are imposed on the instability growth rate.[5] Even for a monoenergetic beam it turns out that $\gamma<\gamma_{\max}$, since the fluctuations of the beam velocity can be significant in a stationary nonlinear wave. Within an order of magnitude $\gamma\sim\Omega_e(n_b/n_0)(V_0/(\Delta V))^2$ (Ref. 4). When $(V_0/(\Delta V))^2\sim(n_0/n_b)^{1/2}$, we find $\gamma\sim\Omega_e(n_b/n_0)^{1/2}$ [see Eq. (15)]. This value corresponds to the minimum instability growth rate excited by a monoenergetic beam.

For a further analysis of the evolution of the system, in (4) we go over to new variables, i.e., amplitude-phase variables, by setting $a_j=A_j\exp(i\varphi_j)$. In the new variables the system (4) acquires the following form:

$$\begin{aligned} &\dot{A}_1=+\lambda_1A_2\sin\Delta\varphi,\quad \Delta\varphi=\varphi_2-\varphi_1,\\ &\dot{A}_2=+\lambda_2A_1\sin\Delta\varphi,\\ &\Delta\dot{\varphi}=-((\alpha_1+\beta_2)A_1^2+(\alpha_2+\beta_1)A_2^2)\\ &\qquad+(\lambda_1A_2/A_1+\lambda_2A_1/A_2)\cos\Delta\varphi. \end{aligned}\tag{6}$$

From the first two equations in (6) we find the integral of motion

$$\lambda_1A_2^2-\lambda_2A_1^2=C.\tag{7}$$

We henceforth assume that the integration constant is equal to zero. Using (7), we obtain from (6) an equation which describes the variation of the phase difference between the waves,

$$\ddot{\psi}+\omega_0^2\sin\psi=0,\quad \psi=2\Delta\varphi,\quad \omega_0^2=4\lambda_1\lambda_2,\tag{8}$$

and an equation which describes the dynamics of the wave energy,

$$\begin{aligned} &\dot{\psi}=-\alpha A_1^2+\omega_0\sin(\psi/2),\\ &\alpha=(\alpha_1+\beta_2)+\lambda_2\lambda_1^{-1}(\alpha_2+\beta_1). \end{aligned}\tag{9}$$

The coefficients appearing in Eqs. (8) and (9) were previously defined in (5).

Equation (8) has the form of the stationary sine-Gordon equation. Its solutions are known and are expressed in terms of the elliptic functions

$$\dot{\psi}(t')=2\varkappa\omega_0\begin{cases}\mathrm{cn}(t',\varkappa), & \varkappa\leqslant 1,\\ \mathrm{dn}(t',1/\varkappa), & \varkappa\geqslant 1,\end{cases}\tag{10}$$

where we have introduced the notation

$$\begin{aligned} &H=\frac{1}{2}\dot{\psi}^2-\omega_0^2\cos\psi,\quad H_s=\omega_0^2,\\ &\varkappa^2=\frac{1}{2}\left(1+\frac{H}{H_s}\right),\quad N=\frac{\omega_0}{\omega(H)}. \end{aligned}\tag{11}$$

Here $\omega(H)$ is the frequency of the nonlinear oscillations, and ω_0 is the frequency of the small oscillations (we note that ω_0 is specified by the linear growth rate of the problem). When $\varkappa=1$, expression (10) goes over to the definition of ψ at the separatrix. Assuming that there is a developed turbulence regime in the system at $t'=0$ and setting $\psi=0$ at $t'=0$ or $(\partial\psi/\partial t')=2\omega_0$ at $\psi=0$, we obtain

$$\dot{\psi}=2\omega_0\cosh^{-1}(\omega_0\,t').\tag{12}$$

When other conditions on $(\partial\psi/\partial t')$ and ψ are chosen, the initial instant of "time" t' is simply shifted, and a constant t_0 must be added to (12) [for example, $t_0=\infty$ for $(\partial\psi/\partial t')=0$ and $\psi=\pi$].

Substituting (10) and (12) into (9), we obtain the expressions for the dynamics of A_1

$$A_1^2=\frac{\omega_0}{\alpha}\begin{cases}\mathrm{dn}(t',\varkappa)+\varkappa\,\mathrm{cn}(t',\varkappa), & \varkappa\leqslant 1,\\ \varkappa\mathrm{dn}(t',1/\varkappa)+\mathrm{cn}(t',1/\varkappa), & \varkappa\geqslant 1,\end{cases}\tag{13}$$

$$A_1^2=\frac{2\omega_0}{\alpha}\cosh^{-1}(\omega_0\,t-k_0x),\quad \varkappa=1.\tag{14}$$

Using the definitions (5), (8), and (9), we can calculate the amplitude value A_1

$$\langle A_1^2\rangle=A_0^2=0.4\left(\frac{n_b}{n_0}\right).\tag{15}$$

Equations (13) and (14) describe the evolution of the energy of nonlinear Langmuir waves. In particular, Eq. (14) corresponds to two types of waves: solitons,

$$A_1=+\left(\frac{2\omega_0}{\alpha}\right)^{1/2}\cosh^{-1/2}\left(\omega_0\left(t-\frac{x}{V_0}\right)\right),$$

and antisolitons (cavitons)

$$A_1=-\left(\frac{2\omega_0}{\alpha}\right)^{1/2}\cosh^{-1/2}\left(\omega_0\left(t-\frac{x}{V_0}\right)\right).$$

Interesting information on the dynamics of the waves can be obtained if we consider the expression for the phase difference when $H=H_s$. Under the initial condition $\Delta\varphi=0$ at $t'=0$, from Eq. (8) we find

$$\Delta\varphi=2\arctan[\exp(\omega_0\, t')-\pi/2]. \tag{16}$$

As $t'\to\infty$, the phases of the waves go out of resonance, and $\Delta\varphi\to\pi/2$.

STOCHASTIC DYNAMICS OF PARTICLES

Let us turn to the spectral properties of the system. It was shown in Ref. 6 that near the separatrix ($\varkappa^2\to 1$, $H\to H_s$) the parameter N, which is defined in (11), has the form

$$N\sim\frac{1}{\pi}\ln\frac{32H_s}{H_s-H}. \tag{17}$$

Near the separatrix $\omega(H)\to 0$, and the oscillation period diverges logarithmically. The phase velocity $\partial\psi/\partial t'$ and the energy $A^2(t')$ of the waves approach a periodic sequence of soliton-like pulses with a distance between crests along the time scale equal to $2\pi/\omega(H)$ and a crest width close to $2\pi/\omega_0$. The spectrum of these modes becomes broad and contains harmonics with the amplitudes[6]

$$b_n\sim 8\omega\begin{cases}1, & N\geqslant n>1,\\ \exp(-n/N), & n>N,\end{cases} \tag{18}$$

i.e., all the harmonics are approximately equal up to $n\sim N$ and are exponentially small at $n>N$. The number N specifies the off-on ratio of the functions $\partial\psi/\partial t'$ and A^2 and the characteristic number of harmonics in the spectrum. As the separatrix is approached, $N\to\infty$, and the spectrum tends to a continuous spectrum.

Let us determine the correlation properties of the system at the separatrix. For this purpose we introduce the correlator $q_1(\tau)=\langle A_1(t')A_1(t'+\tau)\rangle$, where the angle brackets $\langle\ldots\rangle$ denote averaging over the ensemble (time).[6] The spectral power density $q_1(\omega)$ is related to the correlator $q_1(\tau)$ by the expression

$$q_1(\omega)=\int_{-\infty}^{\infty}d\tau\, e^{i\omega\tau}q_1(\tau).$$

Substituting the expressions for $A_1(t')$ from (14) into these definitions, we have

$$q_1(\tau)=\frac{2\pi}{\alpha}\cosh^{-1}(\omega_0\tau),$$

$$q_1(\omega)=\frac{\pi^2}{\alpha\omega_0}\cosh^{-1}(\pi\omega/2\omega_0). \tag{19}$$

It can be seen from (2) and (4) that the wave spectrum is continuous (the characteristic width of the spectrum $\Delta\omega\approx\omega_0$), the decorrelation time $\tau_c=(\omega_0)^{-1}$, and the correlator $q(\tau)$ behaves as $\exp(-\omega_0\tau)$ at $\tau\to\infty$.

Let us investigate the dynamics of electrons in a plasma field. As follows from the foregoing, a high-frequency wave with a slowly varying amplitude, or, stated differently, a broad wave packet, appears in the plasma. We shall show that the trajectories of particles in the field of such a packet become stochastic under certain conditions The phase velocities of the packet tightly fill a certain interval $(\omega/k)_{\max}>V_0>(\omega/k)_{\min}$, so that an effective interaction (Landau resonance) of the waves with the particles occurs in this region. Because of the overlap of resonances, the trajectories of resonant particles become complicated, and regions with stochastic particle dynamics appear on the $(\dot{x},x)$ phase plane. Since the main nonlinear effect is associated with the reverse influence of the plasma oscillations on the distribution of resonant particles,[4] the equation describing the dynamics of the particles will have the following form:

$$\ddot{x}=-\frac{e}{m_e}\sum_k E_k\exp(ikx-i\omega_0 t). \tag{20}$$

Here the Fourier harmonic E_k of the electric field is determined from Poisson's equation and Eqs. (14) and (15). It follows from (17)–(19) that the wave spectrum consists of an even number of harmonics with the wave numbers k_n and the frequencies ω_n, which range from $-\infty$ to $+\infty$. It was shown above that the amplitudes of the harmonics are approximately equal up to a certain number. Then the following simplifying assumptions can be made regarding the structure of the packet:

$$k_n=k_0+n\Delta k,\quad \omega_n=\Omega_e+n\Delta\omega,$$

$$k_0=\Omega_e V_0^{-1},\quad A_n=\text{const},\quad A_n^2N=A_0^2, \tag{21}$$

where A_0 is the density-wave amplitude normalized to n_0 and N is the effective number of harmonics.

Furthermore, a large portion of the plasma electrons have velocities $v<V_0$. In that case Eq. (20) reduces to the problem of the motion of a particle in the field of a temporal packet.[6] Using the conditions (15) and (21) and averaging over the high-frequency oscillations, we obtain an equation which describes the slow evolution of a particle in the wave packet field:

$$\ddot{x}=\varepsilon\frac{\Omega_e^2}{k_0}\cos(k_0x-\Omega_e t)\sum_n A_n\cos(n\Delta\omega t)$$

$$=\frac{\Omega_0^2}{k_0}T\cos\Theta(x,t)\sum_{n=-\infty}^{\infty}\delta(t-nT). \tag{22}$$

Here we have introduced the following notation:

$$\Theta(x,t)=k_0x-\Omega_e t,\quad \Omega_0^2=\varepsilon\Omega_e^2A_0N^{-1/2},$$

$$\omega_0=\Omega_e\left(\frac{n_b}{n_0}\right)^{1/2},\quad T=\frac{2\pi}{\Delta\omega}=\frac{2\pi}{\omega_0}N,\quad \varepsilon=N^{-1}, \tag{23}$$

where T is the characteristic time period of the field, $\tau_c=\omega_0^{-1}$ is the decorrelation time, ε is a small averaging parameter, and Ω_0 is the frequency of the small oscillations of a particle in the potential well created by the central harmonic of the wave packet.

It is easy to write a mapping ($\hat{T}$ is the mapping in Ref. 6) which is equivalent to Eq. (22):

$$\bar{v}=v+\frac{K}{k_0T}\cos\Theta,\quad K=\Omega_0^2T^2,$$

$$\bar{\Theta}=\Theta+\omega(\bar{v})T,\quad \omega(v)=k_0v-\Omega_e, \tag{24}$$

where v, Θ, $\bar{v}$, and $\bar{\Theta}$, are the velocities and phases of the particle, respectively, at the times nT and $(n+1)T$.

The mapping (24) is a standard mapping, and when $K\geqslant 1$, the particle trajectories become stochastic, and the corresponding kinetics are described by an equation like the Fokker–Planck–Kolmogorov (FPK) equation. At small particle velocities, at which the packet is temporal, the FPK equation can be written in the divergent form

$$\frac{\partial f(v,t)}{\partial t}=\frac{1}{2}\frac{\partial}{\partial v}D\frac{\partial f(v,t)}{\partial v}. \tag{25}$$

Here $f(v,t)$ is the distribution function of the particles; D is the diffusion coefficient, which is calculated in the usual manner,

$$D=\left\langle\left\langle\frac{(\Delta v)^2}{T}\right\rangle\right\rangle,\quad \Delta v=\bar{v}-v=\frac{K}{k_0T}\cos\Theta,$$

$$D=\frac{1}{2}\frac{K^2}{k_0^2T^3}, \tag{26}$$

and the angle brackets $\langle\dots\rangle$ denote averaging over the phase.

It can be seen from Eqs. (25) and (26) that

$$\langle v^2\rangle=u_0^2+Dt \tag{27}$$

and that stochastic heating takes place.

If the problem (25) would be solved in a certain bounded velocity range in the absence of a flow of particles from that range, then during the characteristic time $t\geqslant\tau_d$, where

$$\tau_d=V_0^2D^{-1} \tag{28}$$

the equilibrium distribution

$$f(v)=\text{const} \tag{29}$$

would be established in that region.

The distribution (29) has the form of a plateau in velocity space.

We use formulas (24), (26), and (28) to calculate K, which characterizes the degree of stochasticity of the system, the diffusion coefficient D, and the characteristic distribution-function relaxation time τ_d. Utilizing formulas (23) in the calculations, we find the following dependences:

$$K=\left(\frac{n_0}{n_b}\right)^{3/4}N^{1/2},$$

$$D=\frac{1}{2\pi}V_0^2\Omega_eN^{-2},\quad \tau_d=2\pi\Omega_eN^2. \tag{30}$$

Formulas (30) contain N, i.e., the effective number of harmonics in the wave packet. For rough estimates of K and D we can set $N\approx n_0/n_b$. Under this condition expressions (30) take on the following form:

$$K=\left(\frac{n_0}{n_b}\right)^{5/4},$$

$$D=\frac{1}{2\pi}V_0^2\Omega_e\left(\frac{n_b}{n_0}\right)^2,\quad \tau_d=2\pi\Omega_e^{-1}\left(\frac{n_0}{n_b}\right)^2. \tag{31}$$

There are several effects that were not taken into account in describing the stochastic dynamics of electrons. First, if a finite decorrelation time τ_c is taken into account in the FPK equation, then Eq. (25) can contain a description of the stochastic acceleration of the particles.[6] Since the distribution function in a plasma–beam system is nonequilibrium, it is possible that the particles are slowed, rather than accelerated, and impart energy to the wave packet. As a result, this leads to modification of the velocity distribution function and establishment of the stationary distribution (29). Another effect is associated with the influence of friction on the mechanism of stochastic heating (the diffusion time along the stochastic trajectories is of the order of the damping time γ^{-1}). In this case it can easily be shown that the maximum energy which the particles can acquire is

$$\left\langle\frac{m_ev^2}{2}\right\rangle_\infty=\frac{m_eV_0^2K^2}{8\gamma\Omega_e^2T^3}.$$

Substituting the expressions for K and T from (23) and (31) into the last formula, we obtain $(m_ev^2)_\infty/m_eV_0^2\sim\gamma\tau_d$.

For electrons moving with velocities $v>V_0$ (on the "tail" of the distribution function), the energy confinement mechanism is associated with the features of the chaotic behavior of the particles. The dynamics of the particles in this case are described by the equation

$$\ddot{x}=\frac{\Omega_0^2}{k_0}L\cos\Theta(x,t)\sum_{n=-\infty}^{\infty}\delta(x-nL), \tag{32}$$

where $L=2\pi V_0N\omega_0^{-1}$ is the spatial period of the field.

The chaotic behavior of the particles occurs only in a bounded velocity range $(V_0,V_{\max})$, where

$$V_{\max}=(V_0L^2\Omega_0^2)^{1/3}. \tag{33}$$

The averaged evolution of the system is derived from the FPK equation, from which it follows that chaotization of the motion of the particles causes the formation of a plateau in energy space on the tail of the distribution function and that the energy of the particles varies on the average according to the law

$$\left\langle\frac{m_ev^2}{2}\right\rangle\sim\frac{m_eV_0^2}{2}+\text{const}\cdot t^{2/3}. \tag{34}$$

This occurs until the energy of the particles reaches the value $\langle m_eV_{\max}^2/2\rangle$. Using the definitions for Ω_0 and $V_{\max}$ from (23) and (33), we find

$$\left\langle\frac{m_eV_{\max}^2}{2}\right\rangle=\frac{m_eV_0^2}{2}\left(\frac{n_0}{n_b}\right)^{5/6}. \tag{35}$$

Let us evaluate the influence of ions on the dynamics of the system. In the experiments in Refs. 1 and 2 the level of ion fluctuations $\delta n_i/n_i \sim 10^{-2}$, i.e., $\delta n_i/n_i \leqslant \delta n_e/n_e$. If the field of ion fluctuations is totally neglected, then the motion of the heavy particles can be described as the result of their interaction with the Langmuir wave electric field envelope. Reasoning as above, we can write the dynamical equation of the ions in the form

$$\ddot{x}=\varepsilon\frac{\Omega_i^2}{k_0}A_0T\cos\Theta(x,t)\sum_{n=-\infty}^{\infty}\delta(t-nT), \qquad (36)$$

where Ω_i is the ion plasma frequency.

It can easily be shown that because of the weak stochastization of the ion motion [the stochastization parameter $K \sim 2\pi(m_e/m_i)(n_0/n_b)^{5/4} \sim 1$] only slow growth of the energy is possible in the system in the field (36):

$$\left\langle\frac{m_i v_i^2}{T_i}\right\rangle = 1+\frac{m_e}{m_i}\frac{T_e}{T_i}\frac{m_e V_0^2}{T_e}\frac{\Omega_e}{2\pi}\left(\frac{n_b}{n_0}\right)^2, \qquad (37)$$

where T_e and T_i are the electron and ion temperatures, and m_i is the mass of an ion.

DISCUSSION OF RESULTS. CONCLUSION

Let us compare the experimental data with the calculated values. In the experiments in Refs. 1 and 2 the typical plasma and beam parameters were as follows: $T_e=3$ eV, $T_i/T_e=0.1$, $n_0=5\times10^{11}$ cm^{-3}, $\nu_e=\Omega_e/2\pi=5\times10^{9}$ GHz, $m_eV_0^2/2T_e=100$, $n_b/n_0 \sim 10^{-2}$, and the beam pulse duration $\tau \sim 10^{-6}$ s.

Let us first determine the characteristic temporal and spatial scales of the problem. Using Eqs. (8), (14), (19), (23), and (31), we find the field pulse time $\tau_c=\omega_0^{-1}=\Omega_e^{-1}(n_0/n_b) \sim 10^{-8}$ s, the pulse repetition period $T \sim \omega_0^{-1}(n_0/n_b) \leqslant 10^{-6}$ s, the modulation frequency $\omega \sim T^{-1} \geqslant 10^6$ s^{-1}, the distribution-function relaxation time $\tau_d \sim \nu_e^{-1}(n_0/n_b)^2 \sim 10^{-6}$ s, the characteristic width of the potential well created by the central harmonic $l_0 \sim k_0^{-1} = V_0\Omega_e^{-1}=(V_0/V_T)r_d$ (r_d is the Debye radius), and the harmonic interaction scale (the width of the soliton-like pulse) $l \sim V_0\omega_0^{-1} \approx l_0(n_0/n_b)^{1/2}$. It follows from (16) that the turbulence energy level $(\delta n_e/n_0)^2 \approx 0.4$, $(n_b/n_0)^{1/2} \approx 4\times10^{-2}$, and $|E|^2/4\pi n_0T_e \sim 10$. Using formulas (27), (31), and (34) we calculate the electron energy increment during a beam pulse $(m_eV_0^2/2)(n_b/n_0)^2\nu_e\tau \sim 10^{-1}$, $(m_eV_0^2/2)$, and the maximum electron energy on the tail of the distribution function $(m_eV_0^2/2)(n_0/n_b)^{5/6}$. An estimate for ions from (37) shows that the relative ion energy increment $(m_iv_i^2/T_i) \sim (m_e/m_i)(T_e/T_i)(m_eV_0^2/T_e)(n_b/n_0)^2\nu_e\tau \sim 0.1$.

In summation, it has been shown in this paper that the induced scattering of beam electrons excites nonlinear Langmuir waves in the plasma. The electric field envelope of the Langmuir oscillations has the form of a cnoidal wave. The dynamics of electrons in the wave field becomes chaotic, and the stochasticity is manifested in several macroscopic effects. The results obtained within this model provide explanations for several experimentally observed features of the evolution of a plasma–beam system.

[1] D. A. Whelan and R. L. Stenzel, Phys. Rev. Lett. **47**, 95 (1981).

[2] D. M. Karfidov, A. M. Rubenchik, K. F. Sergeĭchev, and I. A. Sychev, Zh. Éksp. Teor. Fiz. **98**, 1592 (1990) [Sov. Phys. JETP **71**, 892 (1990)].

[3] T. Neubert and P. M. Banks, Planet. Space Sci. **39**, 1 (1991).

[4] L. A. Artsimovich and R. Z. Sagdeev, *Plasma Physics for Physicists* [in Russian], Atomizdat, Moscow (1979).

[5] V. L. Ginzburg, *Applications of Electrodynamics in Theoretical Physics and Astrophysics*, 2nd rev. ed., Gordon and Breach, New York (1989).

[6] G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics* [in Russian], Nauka, Moscow (1988).

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Structure of group cathode spots on the surfaces of hot-rolled steels

S. L. Pozharov, A. M. Mirkarimov, and I. V. Soldatov

Institute of Electronics, Uzbekistan Academy of Sciences, 700143 Tashkent, Uzbekistan

(Submitted April 29, 1997)
Zh. Tekh. Fiz. **68**, 57–62 (November 1998)

The results of investigations of cathode spots on steel surfaces are presented. Their behavior is found to have some distinctive features in comparison with the results obtained on clean metals. © *1998 American Institute of Physics.* [S1063-7842(98)01111-8]

INTRODUCTION

Research on the dynamics of cathode spots on various surfaces is of interest largely in connection with the solution of practical problems, particularly with the development of technologies for the vacuum-arc treatment of the surface of electrically conductive materials.[1–3] Most of the known experimental results have been obtained in studies performed on clean metal surfaces.[4–7] At the same time, for practical purposes it is important to know the features of the behavior of cathode spots on surfaces covered by various films, particularly fairly thick layers of scale. Such studies have not been carried out to any appreciable extent on account of the complexity of the interpretation of the results and their poor reproducibility, which is due to the nonuniformity of oxidized surfaces.

The preliminary results of investigations of cathode spots on the surface of hot-rolled steel performed by the ''autograph'' method[4] were recently reported in Ref. 8. It was established that elementary cathode spots, which are also known as cells[4] or as independent or separate cathode spots,[5] exist on such a surface in the form of compact groups, which determine the type of erosion of the surface. A theory on such compact groups of elementary cathode spots was previously advanced in Ref. 5, where they were termed group cathode spots.

The purpose of this paper is to report the results of further, more systematic investigations of group cathode spots. Samples of hot-rolled stainless steels 304 and 430 were selected as objects of investigation.

EXPERIMENTAL METHOD

The idea behind the investigation of group cathode spots by the ''autograph'' method is exceptionally simple[4] and can be described as follows. As a group cathode spot moves over a surface, it leaves an imprint, which forms as a result of erosion of the surface. If the surface is covered by an oxide layer (scale), its erosion includes removal of the scale from it, as a result of which the imprint formed has a very clearly outlined contour. Measuring the geometric parameters of the imprint (length, width, and area) for a fixed discharge burning time permits determination of the dimensions of the group cathode spot and its mean rate of motion over the surface. The geometric parameters of imprints were measured on an automated image-processing system with input of the images into a computer from a television camera with an assigned magnification.

A pulsed discharge power supply, which generates pulses of rectangular shape with a regulatable duration $T=1-1000$ ms and an amplitude up to $I=300$ A, was used.

The following discharge parameters were measured independently: the discharge current I, the voltage on the discharge gap U, the total electric charge passing through the discharge during a single pulse $Q=\int_0^T I(t)\,dt$, and the discharge pulse duration T.

The electrode system had a flat geometry with an electrode gap much smaller than the linear dimensions of the electrodes, and the sample being investigated served as the cathode. The experimental setup permitted the performance of investigations both in the absence of an external magnetic field and in an external magnetic field.

In the former case a cassette holder for six samples, whose geometry ensured the most symmetric distribution of the local magnetic fields created by the discharge current and the leads over the surface of each sample, was mounted in the vacuum chamber. Investigations associated with determination of the most probable value of the current per group spot, as well as other parameters of the group cathode spots under conditions allowing their free movement over the surface, were performed in this variant.

In the latter case an electromagnet with one test sample in its opening was mounted in the vacuum chamber. The maximum magnetic field strength over the surface of a nonmagnetic material was $H=1.3\times10^5$ A/m. The corresponding value over the surface of a sample of a material having ferromagnetic properties was $H=8.6\times10^4$ A/m. The dependence of the width of the imprint of a group cathode spot (its dimensions) and the rate of motion on discharge current, as well as the influence of the magnetic field on the properties of the group cathode spot, were investigated in this variant.

The value of Q was measured by a special instrument to a relative error of 2%. The automated image-processing system enabled us to measure the area of the cleaned surface to a relative error of 5% and the linear geometric parameters of the imprints of group cathode spots to a relative error of 3.5%. The duration of the discharge pulses was varied with 1-ms steps and established to within ±10 μs.

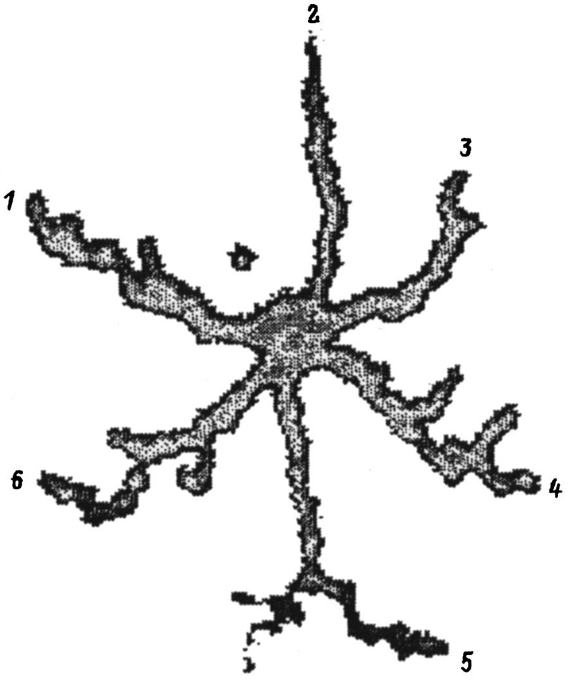

FIG. 1. ''Autograph'' of group cathode spots (*1–6*) on a surface of steel 430 for $I=150$ A and $T=0.015$ s.

EXPERIMENTAL RESULTS

In the absence of an external magnetic field at sufficiently large discharge currents (>100 A) the ''autograph'' of cathode spots on a surface (the outline of the area freed of scale) has a clearly expressed structure in the form of almost symmetrically dispersed, winding tracks of irregular shape, on which there are branches. A typical example of an ''autograph'' is shown in Fig. 1. At small currents, for example, at $I=20$ A, such a structure is not observed even for large discharge-current durations: the ''autograph'' has the form of a spot of irregular shape. One interesting feature of the microrelief of an eroded surface is the presence of axially symmetric columnar protrusions of the fused metal, which reach heights up to several tens of microns. They are positioned randomly at the center of the ''autograph'' and on widened portions of the tracks. On narrow portions of well formed tracks the protrusions are located predominantly along the axial line of the track, sometimes fusing and forming a structure resembling an axial ridge. This is displayed especially clearly on the tracks formed in the presence of an external magnetic field, which will be discussed below.

The measurements were performed for various amplitudes ($110-160$ A) and durations ($5-50$ ms) of the discharge current. The most probable value of the current i_g in a group cathode spot was determined by counting the number of tracks N_k at each value of the discharge current under the assumption that all the tracks form simultaneously. It must be admitted that the procedure for counting the number of tracks is very subjective. Therefore, several counting criteria were used: for example, only tracks diverging from the center were counted (without consideration of branching); short branches were not counted; and allowance was made for the possibility of two group cathode spots moving along a track simultaneously, which would be displayed in the form of one highly branched track. It turned out that the result of calculating i_g depends weakly on the choice of the criterion for counting N_k. The values of i_g determined according to different criteria, were confined to the following ranges: $i_g=25.1-28.7$ A for steel 430 and $i_g=22.5-23.7$ A for steel 304. The best criterion for determining N_k was estimated from the minimum of the quantity

$$\Delta=\sum_{k=1}^{n}(N_k-I_k/i_g)^2/(n-1),$$

where n is the number of ''autographs'' treated, I_k is the discharge current for the formation of the kth ''autograph,'' and N_k is the number of tracks counted in the kth ''autograph.''

The following values were taken as a result: $i_g=27.5\pm0.7$ A for steel 430 and $i_g=23.7\pm1.0$ A for steel 304. Under these conditions the variance of the normal distribution of the current in a group cathode spot is $D(i_g)=6.8$ A^2 for steel 430 and $D(i_g)=15.2$ A^2 for steel 304.

The mean values of the width (d) of the tracks of group cathode spots were determined. The width of each track was measured at 10 equidistant points. The results of many measurements performed for different discharge pulse amplitudes and durations were treated according to the least-squares method. No correlation was discovered between the track width and the discharge current (the number of tracks for the group cathode spot), as well as the discharge pulse duration. The following values were obtained: $d=(0.045\pm0.006)$ cm with a variance $D(d)=0.0002$ cm^2 for steel 430 and $d=(0.084\pm0.008)$ cm with a variance $D(d)=0.0005$ cm^2 for steel 304.

It is easy to see that the distribution function of the group cathode spots with respect to the relative value of the current flowing in them is narrower than their distribution with respect to the measured relative track width. Thus, in our opinion, the great spread of measured track widths is determined to a considerable extent by the superposition of the chaotic component of the velocity of the group cathode spots on the radial component. This is also supported by the character of the surface microrelief in the regions of appreciable widening of the tracks (the chaotic arrangement of the protrusions on the surface).

The radial components of the velocity of group cathode spots were measured. Anisotropy of the velocity of the group cathode spots on the surfaces of the samples investigated was discovered. The velocity of the group cathode spots in the rolling direction was higher than the velocity in the transverse direction. The radial components of the velocity of the group cathode spots decrease with increasing distance from the discharge initiation center, in agreement with the conception of the influence of the total magnetic field created by group cathode spots on their motion. Plots of the dependence of the radial component of the velocity in the rolling direction ($V_{\max}$) and in the transverse direction ($V_{\min}$) on the discharge pulse duration for both types of steel are shown in Fig. 2. Processing of the results of measurements of area freed of scale (S) for various values of $Q=\int_0^T I(t)\,dt$ showed

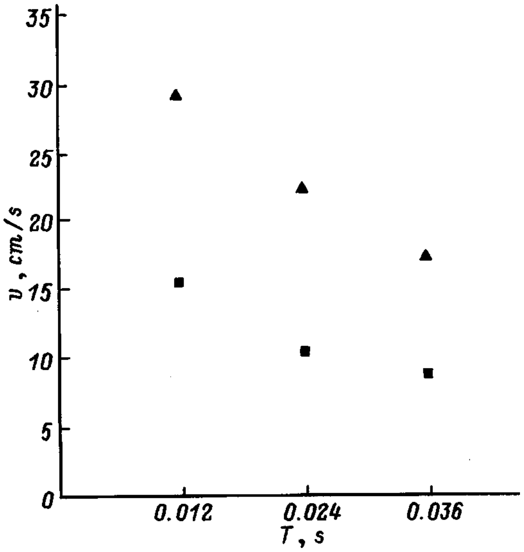

FIG. 2. Dependence of the radial component of the velocity of group cathode spots on the duration of the discharge pulse for steel 304: ■ — $v_{\min}$, ▲ — $v_{\max}$.

that the relationship between these quantities at a fixed current and pressure is approximated well by the linear dependence $S=Q/q_{\text{eff}}$ (the linear correlation coefficient is no poorer than 0.99). Thus, for fixed current and pressure values q_{eff} is a constant (which does not depend on time). The constancy of q_{eff} with time attests to the constancy of the fraction of the current in a group cathode spot which removes the scale from the surface.

For steel 430 we have $q_{\text{eff}}=(11.8\pm 1.5)$ C/cm^2, and for steel 304 we have $q_{\text{eff}}=(13.8\pm 0.6)$ C/cm^2. The parameter q_{eff} has a weakly expressed maximum in the range of discharge currents $140-150$ A, remaining essentially constant for other values of the current.

An external magnetic field was used to induce the directed motion of a single group cathode spot for the purpose of subsequently determining the dependence of the rate of motion and the characteristic dimensions of the group cathode spot on the current value of the current in it. A group cathode spot moves in an external magnetic field in the direction opposite to the Ampère force.

Measurements were performed for three values of the magnetic field and 6–10 values of the discharge current. The minimum value of the magnetic field strength was chosen so that a straight track of approximately invariant width with a regular microrelief in the form of an axial ridge consisting of closely arranged and partially merging protrusions would form.

The following values of the magnetic field strength H were chosen: $H_l=2.23\times 10^4$ A/m, $H_m=4.5\times 10^4$ A/m, and $H_h=6.75\times 10^4$ A/m for steel 430; $H_l=2.4\times 10^4$ A/m, $H_m=4.7\times 10^4$ A/m, and $H_h=8\times 10^4$ A/m for steel 304. The values of the discharge currents were chosen in the range $i_g/2<I<2i_g$, which corresponds to the existence region of a single group cathode spot. The velocity of the group cathode spot in these measurements was found as $v=l/T$, where l is the track length and T is the discharge pulse duration.

Two types of dependences of the mean track width d on the discharge current I were discovered: a linear dependence and a dependence of the form $d=k_d\sqrt{I}$. The dependence of d on I for steel 430 turned out to be nearly linear at all magnetic field strengths. The dependence of d on I for steel 304 is nearly linear at low magnetic fields (H_l), while at high fields (H_h) it has the form $d=k_d\sqrt{I}$. With respect to the dependence at medium values of the magnetic field (H_m), the spread observed in the experimental data precludes making an unequivocal choice between these two types of dependences.

The results of the measurements of the velocity of group cathode spots as a function of discharge current showed that a regular dependence is observed in only one case, viz., for steel 304 at a high magnetic field, and has the form $v=k_v\sqrt{I}$. In the remaining cases it can be stated that their is no dependence of the velocity of group cathode spots on current.

When the dependence of d on I is linear, the characteristic track microrelief (the axial ridge consisting of merging protrusions) does not undergo any qualitative changes. The dimensions of the axial ridge decrease as the track width decreases. The transition to the dependence of the form $d=k_d\sqrt{I}$ (steel 304 at a high magnetic field) is accompanied by qualitative changes in the microrelief. More specifically, the pronounced axial ridge disappears, and an appreciable quantity of fused scale appears on the bottom of the track.

The measurements of $q_{\text{eff}}=Q/S$, where S is the track area, showed that it does not depend on discharge current. The value of q_{eff} for steel 430 does not depend on magnetic field strength to within the measurement accuracy range and remains the same as in the absence of an external magnetic field. At the same time, for steel 304 there is an appreciable decrease in q_{eff} in an external magnetic field, the strongest change occurring upon the transition from fields of medium strength to fields of high strength. The values of q_{eff} for various magnetic field strengths are listed in Table I.

To complete this section we note that a significant contribution to the statistical spread of the measured parameters of group cathode spots could be made by the inhomogeneity of the scale on the sample surface. In order to clear up this question, the thickness of the scale on a large number of samples (22 samples for each kind of steel) was measured. The results were as follows: a mean scale thickness of 7.0 μm with a variance of the normal distribution of the thickness equal to 2 μm^2 for steel 430, and a mean scale thickness of 7.6 μm with a variance of the thickness distribution equal to 0.7 μm^2 for steel 304.

DISCUSSION OF RESULTS

Examining the results obtained, we start out from the most widely used conceptions of an elementary cathode spot as a local active emission center, which provides for the flow of large currents with a comparatively small voltage on the discharge gap. According to these conceptions, an elementary cathode spot is a highly nonstationary and unstable system, which forms, develops, and ends its existence in a fixed local region on the surface. The movement of an elementary cathode spot over the surface should be regarded as the result

TABLE I. Types of dependences of the parameters of group cathode spots on the discharge current, their configuration, and the value of q_{eff} for various external magnetic field strengths.

Magnetic field strength	Type of steel	304			430		
	Parameters	d, cm	v, cm/s	q_{eff}, C/cm^2	d, cm	v, cm/s	q_{eff}, C/cm^2
Low (H_l)	Type of dependence	$d=k_dI$ $k_d=(1.92\pm0.05)\times10^{-3}$	$v=\mathrm{const}$ $v=62\pm4$	8.1 ± 0.5	$d=k_dI$ $k_d=(1.71\pm0.17)\times10^{-3}$	$v=\mathrm{const}$ $v=46\pm7$	12.8 ± 1.0
	Configuration of group cathode spot	One-dimensional			One-dimensional		
Medium (H_m)	Type of dependence	$d=k_dI$ $k_d=(1.31\pm0.12)\times10^{-3}$	$v=\mathrm{const}$ $v=99\pm6$	7.4 ± 1.0	$d=k_dI$ $k_d=(1.01\pm0.10)\times10^{-3}$	$v=\mathrm{const}$ $v=84\pm10$	11.1 ± 1.1
	Configuration of group cathode spot	One-dimensional			One-dimensional		
High (H_h)	Type of dependence	$d=k_d\sqrt{I}$ $k_d=(5.93\pm0.23)\times10^{-3}$	$v=k_v\sqrt{I}$ $k_v=26.1\pm0.8$	5.6 ± 0.5	$d=k_dI$ $k_d=(0.98\pm0.06)\times10^{-3}$	$v=\mathrm{const}$ $v=94\pm7$	10.2 ± 0.8
	Configuration of group cathode spot	Two-dimensional			One-dimensional		

of the successive formation of new elementary cathode spots in the immediate vicinity of disappearing elementary spots. The characteristic parameters of elementary cathode spots, such as their probability (frequency) of formation, lifetime, linear dimensions, and mean current, depend both on the physical properties of the surface in the region where they form and on the initial local conditions (the density of the near-surface plasma and the temperature) and do not depend on total discharge current, i.e., the discharge current is directly proportional to the number of elementary spots on the cathode.

Unlike an elementary cathode spot, a group cathode spot is a stable system, whose existence is sustained by the equilibrium between the rate of formation and the rate of destruction of the elementary cathode spots comprising it. The movement of a group cathode spot over the surface is most likely attributable to the difference in the probability of the formation of elementary cathode spots in certain selected directions due to the inhomogeneous distribution of the near-surface plasma density in local magnetic fields.

The following qualitative explanation for the formation of a group cathode spot on a surface covered by scale can be conjectured. An isolated elementary cathode spot on such a surface obviously cannot provide the necessary initial conditions for the formation of a new elementary cathode spot with a sufficiently high probability and thus cannot create successively appearing new elementary cathode spots over the course of any prolonged time period. If there are several fairly closely arranged elementary cathode spots on the surface, their combined influence on the surface results in the creation of more favorable conditions for the formation of new elementary cathode spots to replace the disappearing spots. Thus, a group of elementary cathode spots is a more viable system than an isolated elementary cathode spot. In addition, the fact that each type of surface has a definite, most probable value of the current in a group cathode spot (a most probable number of elementary cathode spots comprising a group cathode spot) with a comparatively small variance seems to be evidence of the operation of a mechanism which displaces the equilibrium between the rate of formation and the rate of disappearance of elementary cathode spots in accordance with their number with a group cathode spot. When the current in a group cathode spot (the number of elementary cathode spots) is less than the most probable value, the rate of formation of elementary cathode spots is higher than the rate of disappearance. In the opposite case the rate of formation becomes less than the rate of disappearance.

The results of the measurements of the dependence of the group cathode spot track width on the current in an external magnetic field showed that in all cases, with the exception of steel 304 in a high external magnetic field, a group cathode spot has a configuration in which the elementary cathode spots are arranged along the outline of the group spot according to a law which does not depend on current. In other words, the mean distance a between the elementary cathode spots remains constant as the current in the group spot varies. Below we shall call such a configuration one-dimensional by convention. The extreme examples of one-dimensional configurations of group cathode spots are a linear segment of length $\sim d$ and a circle of diameter $\sim d$. Here

$$d=k_dI, \tag{1}$$

where $k_d=a/fi_e$, i_e is the mean current in an elementary cathode spot, and f is the geometric factor.

The dependence of the form

$$d=k_d\sqrt{I} \tag{2}$$

obtained for steel 304 in a high magnetic field corresponds to a configuration of the group cathode spot in which the elementary spots are distributed over the entire area of the group spot according to a law which does not depend on current. Below we shall call such a configuration two-dimensional. In this case

$$k_d = a/\sqrt{i_e},$$

where a is the mean distance between elementary cathode spots, which likewise does not depend on the current in the group cathode spot.

The conceptions of group cathode spots with one- and two-dimensional configurations are in good agreement with the results of the measurements of the velocity of group cathode spots as a function of current.

Let us consider the formation of a surface freed of scale in the form of a straight track as a result of the motion of a group cathode spot. Then the area of the cleaned surface is specified by the expression

$$S = dvt, \tag{3}$$

where v is the rate of motion of the group cathode spot.

On the other hand, the cleaning of the surface can be regarded as the result of the interaction of the elementary cathode spots forming a group spot with the scale. In this case

$$S = \chi n \frac{s_e}{\tau_e} t, \tag{4}$$

where n is the number of elementary cathode spots forming the group spot; χ is the effective fraction of elementary cathode spots interacting with the scale $(\chi \leqslant 1)$, s_e is the area of the cleaned surface formed during the lifetime of an elementary cathode spot as a result of its interaction with the scale, and τ_e is the lifetime of an elementary cathode spot.

Comparing the right-hand sides of (3) and (4), we find

$$v = n \frac{1}{d} \chi \frac{s_e}{\tau_e} = \frac{I}{i_e} \frac{1}{d} \chi \frac{s_e}{\tau_e}.$$

Using formulas (1) and (2), we can easily find the dependence of the velocity of a group cathode spot on current for different structures:

$$v = \begin{cases} \text{const} = \dfrac{f}{a} \chi \dfrac{s_e}{\tau_e} & \text{for a one-dimensional configuration,} \\[2ex] k_v \sqrt{I}, \ \text{where } k_v = \dfrac{1}{a\sqrt{i_e}} \chi \dfrac{s_e}{\tau_e} & \text{for a two-dimensional configuration.} \end{cases}$$

The types of the dependences of the parameters of a group cathode spot and its structure on discharge current are given in Table I. It was assumed in all the arguments above that the basic parameters of elementary cathode spots do not depend on discharge current.

The observed axially symmetric structure of typical elements of the microrelief of an eroded surface, i.e., the protrusions, leads to an intuitive picture of the analogous symmetry in a group cathode spot. For example, in the case of a one-dimensional configuration it suggests a group cathode spot of annular structure with a uniform distribution of elementary cathode spots around a circle. However, the movement of the group cathode spot over the surface is at variance with this intuitive picture. As we have already stated above, the motion of a group cathode spot is caused by the higher probability of the formation of new elementary cathode spots in the direction of motion of the group spot, and, therefore, there cannot be a uniform distribution of elementary cathode spots along the outline or over the area of the group spot. It can theorized, for example, that an overwhelming majority of elementary cathode spots in the region of a group cathode spot with a one-dimensional configuration will be concentrated on the arc of a circle facing the direction of motion of the group cathode spot at a certain value of the magnetic field strength. The elementary structure of the microrelief of an eroded surface is displayed most clearly when the duration of the discharge pulse is small in the absence of a magnetic field. At durations less than 1 ms an eroded surface has the form of individual, closely arranged craters with a nearly circular outline. The following crater formation dynamics are observed. In the initial stage of development of the discharge the dimensions of the craters grow with increasing time. The bottom of each crater has a flat shape at its center with a considerable quantity of fused scale and islands of clean metal. There are subsequently qualitative changes in the microrelief of the crater without significant changes in its dimensions. A columnar protrusion of fused metal appears at the center of the crater, and a clearly expressed border of fused scale forms along the edges. A schematic representation of such a crater is shown in Fig. 3.

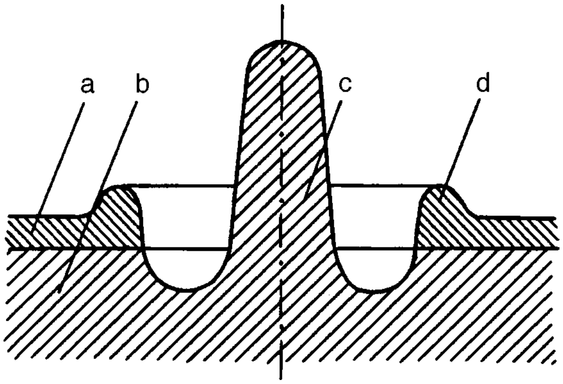

FIG. 3. Schematic representation of a crater (in section) in the final stage of the formation of a typical microrelief: a — scale layer, b — metal, c — protrusion, d — border of fused scale.

Approximate estimates of the time of formation of the characteristic microrelief were obtained: $50-100\ \mu$s for steel 430, and $100-150\ \mu$s for steel 304. In our opinion, the processes which shape the structure of a group cathode spot and the microrelief of a crater are directly related to one another. In this situation, a group cathode spot apparently has a two-dimensional configuration at the initial moment, and then it relaxes to a one-dimensional configuration during a finite time.

The fact that passage from a one-dimensional to a two-dimensional configuration of a group cathode spot was observed on steel 304 at high magnetic fields in the experiments in an external magnetic field is confirmed not only by the dependence of v and d on the discharge current, but also by the changes observed in the track microrelief in this case.

It is significant that the microrelief of the track left on a surface by a group cathode spot of two-dimensional configuration is qualitatively similar to the microrelief of the craters formed in the initial stage of discharge development. The significant decrease in the value of q_{eff} upon passage from the one- to the two-dimensional configuration agrees well with the qualitative change in the microrelief of the cleaned surface, since an appreciable portion of the energy expended on cleaning the surface must be used to form the protrusions.

The absence of a change in the structure of a group cathode spot on steel 430 is probably attributable to the shorter relaxation time of a group cathode spot and the lower rate of motion of a group cathode spot at high magnetic fields on this type of steel than on steel 304.

CONCLUSIONS

1. The elementary cathode spots on the surfaces of hot-rolled steels 304 and 430 in a quasistationary vacuum arc discharge at large currents exist in the form of compact structurally ordered groups, i.e., group cathode spots. The properties of group cathode spots are characterized by the mean value of their diameter, the current, and the rate of their motion over the surface with comparatively small variances.

2. At small discharge currents ($I<2i_g$, where i_g is the mean current of the group cathode spots) a discharge contains one group cathode spot. At large currents ($I>2i_g$) the group cathode spot tends to divide. No qualitative changes in the structure of the group cathode spot are observed as the current is varied. The mean distance between the elementary cathode spots in a group cathode spot and their basic properties do not depend on the current flowing in the group cathode spot.

3. In most of the investigations performed the group cathode spots had a one-dimensional configuration. The formation of group cathode spots with a one-dimensional configuration is unequivocally associated with the formation of a characteristic microrelief on the eroded surface in the form of axially symmetric columnar protrusions of fused metal.

4. Passage from the one-dimensional configuration to the two-dimensional configuration of a group cathode spot is observed on steel 304 at high magnetic fields. The change in the structure of the group cathode spot is accompanied by qualitative changes in the dependences of the width of the track of the group cathode spot and its rate of motion on the current flowing in it, alteration of the microrelief (disappearance of the protrusions) and a decrease in the specific energy expended to free the surface of scale (a decrease in q_{eff}).

In conclusion we would like to express our gratitude to the developers of the image-processing system, T. A. Zakirov, R. D. Sadykov, and E. I. Zyat'kova (Laboratory of Applied Physics, Institute of Electronics, Uzbekistan Academy of Sciences) for their assistance in performing the measurements.

[1] B. E. Bulat, A. M. Mirkarimov, R. B. Nagaibekov *et al.*, in *Proceedings of TATF'96* (1996), pp. 155–157.
[2] U. A. Arifov *et al.*, European Patent No. 0468110 (1990).
[3] S. L. Pozharov *et al.*, Uzbekistan Preliminary Patent No. 600 (1992); Russian Federation Patent No. 93003651/12 (1993).
[4] I. G. Kesaev, *Cathode Processes in an Electric Arc* [in Russian], Nauka, Moscow (1968), 244 pp.
[5] G. A. Lyubimov and V. I. Rakhovskiĭ, Usp. Fiz. Nauk **125**, 665 (1978) [Sov. Phys. Usp. **21**, 693 (1978)].
[6] *Vacuum Arcs. Theory and Applications*, J. M. Lafferty (Ed.), Wiley, New York (1980).
[7] A. I. Bushik, T. A. Bakuto, P. S. Zacherilo, and V. A. Shilov, in *International Symposium on Discharges and Electrical Insulators in Vacuum*, St. Petersburg (1994), p. 96.
[8] A. M. Mirkarimov, S. L. Pozharov, and I. V. Soldatov, Ukr. Fiz. Zh. No. 4, 45 (1996).

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Leader discharge over a water surface in a Lichtenberg figure geometry

V. P. Belosheev

S. I. Vavilov State Optical Institute, All-Russian Scientific Center, 199034 St. Petersburg, Russia

(Submitted September 14, 1997)
Zh. Tekh. Fiz. **68**, 63–66 (November 1998)

A multichannel leader discharge over a water surface is investigated in a Lichtenberg figure geometry. It is established that the Ohmic conductivity of water causes nonlinearity of the $R(t)C$ discharge circuit. A mutual one-to-one correspondence between the channel lengths and the currents flowing in them is established during the discharge, and the discharge has a self-consistent character. A mechanism is proposed for the initiation of initial channels by maxima which arise in the charge structure of the planar double layer on the water surface during the development of Rayleigh–Bénard instability in the layer after the pulsed corona from the anode reaches the water. © *1998 American Institute of Physics.* [S1063-7842(98)01211-2]

In Ref. 1 we investigated the conditions for the formation of a single-channel spark discharge over a water surface with the intention of using it for the UV disinfection of water. The results of that work enabled us, in particular, to disclose the unique features of the evolution of a spark discharge over a water surface due to its conductivity in comparison to the development of a spark discharge over the surface of a solid dielectric.[2] The influence of these features of water on the development of a single-channel leader of a spark discharge over a water surface was investigated in Ref. 3. In that study, on the basis of experimentally determined distributions of the electric field and current density along the leader we found that its development has a self-consistent character and that the product of the storage capacitance and the initial potential difference between the tip of the leader and the water surface beneath it is an invariant of channel development. A natural continuation of the work in Refs. 1–3 would be an investigation of a multichannel leader over water in the geometry used to record Lichtenberg figures of discharges over solid dielectrics. This is the subject of the present paper.

EXPERIMENTAL SETUP AND CONDITIONS

A diagram of the experimental setup is shown in Fig. 1. It includes a cylindrical glass cell *1* with a diameter of 9 cm and a height of 1.5 cm, which is two-thirds full of tap water *2* with a conductivity $\simeq 1\times10^{-4}$ S/cm. A brass disk *3* with a diameter of 7 cm and a thickness of 0.05 cm, which served as the cathode, was positioned concentrically to the cell at a distance of 0.3 centimeters below the surface of the water. The tip of a stainless steel wire of diameter 7.5×10^{-3} cm, which was placed over the center of the cell at a distance of 0.1–0.3 cm from the water surface, served as anode *4*. A storage capacitor *5* with a capacitance of 0.1 μF, which was charged to an initial voltage $U_0=3-6$ kV, was used in the experiments. The discharge over the water surface was initiated by decreasing the width of the air gap in a spark gap *6*. The capacitor voltage U_c and the discharge current i were recorded on an oscillograph during evolution of the discharge using a resistive voltage divider *7* and shunt *8*. The discharge was photographed by camera *9* on Izopankhrom film (type 22) with the same *f* stop in all cases.

EXPERIMENTAL RESULTS

Oscillograms of U_c and i (Fig. 2, curves *1* and *2*) and integral photographs of the discharges (Fig. 3a–3d) were obtained in the experiments in the range $U_0=3-6$ kV. The current oscillograms enable us to determine the amplitudes of the initial current $i_0=0.3$ and 3 A at $U_0=3$ and 6 kV, respectively, and to calculate the initial resistance of the discharge circuit $R_0=U_0/i_0$, and the oscillograms of U_c and i permit determination of the resistance of the discharge circuit at other moments in time (Fig. 2, curve *3*). In addition, the amplitude of the maximum current i_M and the time of its achievement t_M were determined from the current oscillograms, enabling us to calculate the charge transferred up to that moment $q_M\simeq(i_M\cdot t_M)/2$ and to construct plots of the dependence of all these quantities on U_0 (Fig. 4).

Figure 3 shows the central diffuse luminescent region, the channels developing from it, and the image of the wire anode, which is defocused as the distance from the water surface increases (Figs. 3b–3d). In the range $U_0=3-6$ kV the diameter of the diffuse region determined from negatives increases approximately linearly from $\simeq0.15$ to $\simeq0.3$ cm. As in Ref. 3, each channel has side branches of various length, which are directed at an angle to the channel axis, and branches with a length of the order of the channel diameter, which are normal to its axis (Fig. 3a). While branching occurs over virtually the entire length of the channels when $U_0=3$ kV, it occurs only near their beginning when $U_0=6$ kV. The structure of the discharge channels, which is practically symmetric for $U_0=3$ kV, is asymmetric when U_0 is large. This may be due to the nonparallel configuration of the cathode and the water surface. In this case, the wedged shape of the water layer has an appreciably stronger effect on channels of large length.

The number of initial channels for $U_0=3$ kV was 3–6. Out of 29 discharges the probabilities of the appearance of 3 (Fig. 3b), 4 (Fig. 3c), and 5 and 6 (Fig. 3d) initial channels

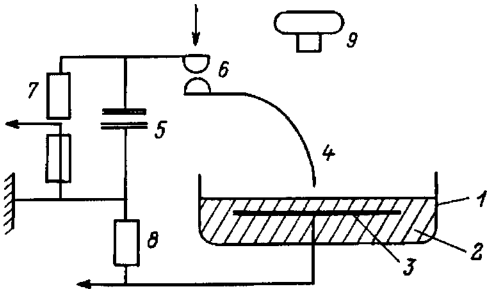

FIG. 1. Diagram of the experimental setup.

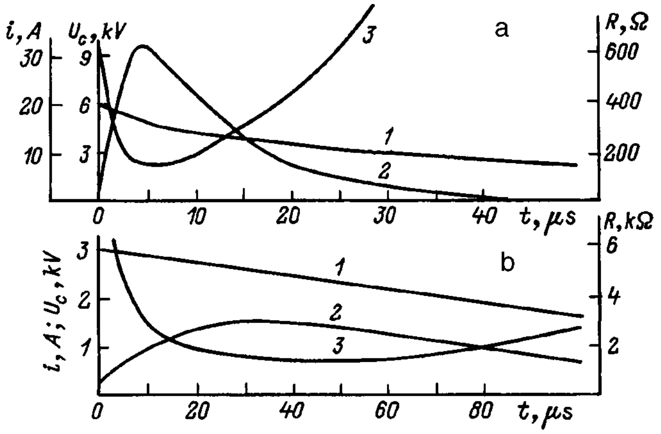

FIG. 2. Oscillograms of the capacitor voltage (*1*) and the discharge current (*2*) and time dependence of the resistance of the discharge circuit (*3*) (U_0, kV: a — 6, b — 3).

are equal to 0.17, 0.62, 0.14, and 0.07, respectively. In the case of $U_0 = 6$ kV the number of initial channels was 3 or 4 with roughly equal probabilities. The radius of the structure or the averaged maximum channel length l_M depends linearly on U_0 (Fig. 4). Using these data and the time of the achievement of the maximum channel length, we can determine the mean velocity of channel development for various values of U_0 (Fig. 5).

DISCUSSION OF THE EXPERIMENTAL RESULTS

The experimentally obtained oscillograms of U_c and i and the integral photographs of a multichannel leader do not permit tracing the dynamics of the development of a discharge and its specific electrical characteristics. This was accomplished under the conditions of a single-channel leader[3] by calling upon data on the distribution of the current and voltage along a discharge channel with time obtained by a probe technique. However, since the conditions for performing the present experiments (the conductivity of the water, the thickness of the water layer above the cathode, and the range of initial voltages) were the same as in Ref. 3 and since the oscillograms of the capacitor voltage and the current were qualitatively similar in both cases, it can be presumed that there is qualitative similarity between the processes involved in the development of single-channel and multichannel leaders. The quantitative differences are associated with the large area of the two-dimensional cathode and the total area of the channels, i.e., the smaller resistance of the discharge circuit in the present cases. Accordingly, on the basis of Refs. 1 and 3 and the present experiments it can be assumed that the presence of a positive voltage on anode *4* gives rise to negative polarization charges on the water surface with an axisymmetric density distribution, in accordance with the distribution of the electric field in the anode–cathode gap. Then a pulsed corona begins to develop from the anode, and a capacitive current appears in the discharge circuit. The duration of this phase of the discharge is $\simeq 0.1$ μs. At the moment when the corona cone touches the water surface, the conductivity of the entire anode–cathode gap takes on an Ohmic character. The amplitude of the current at that moment $i_0 = 0.3$ A for $U_0 = 3$ kV and 3 A for $U_0 = 6$ kV. Such a great difference between the values of the initial current occurs because not only the cross-sectional area of the corona cone on the water surface, but also the cross section of the current-conducting layer of water above the cathode increase when $U_0 = 6$ kV. This conclusion is supported by the fact that $i_0 = 0.2$ A in the case of a one-dimensional wire cathode of diameter 0.16 cm, with $U_0 = 6$ kV and the other conditions as in Ref. 3.

The further increase of the current in the circuit is now associated with the development of discharge channels over the water surface. In this stage there is a mutual one-to-one correspondence between the lengths of these channels and

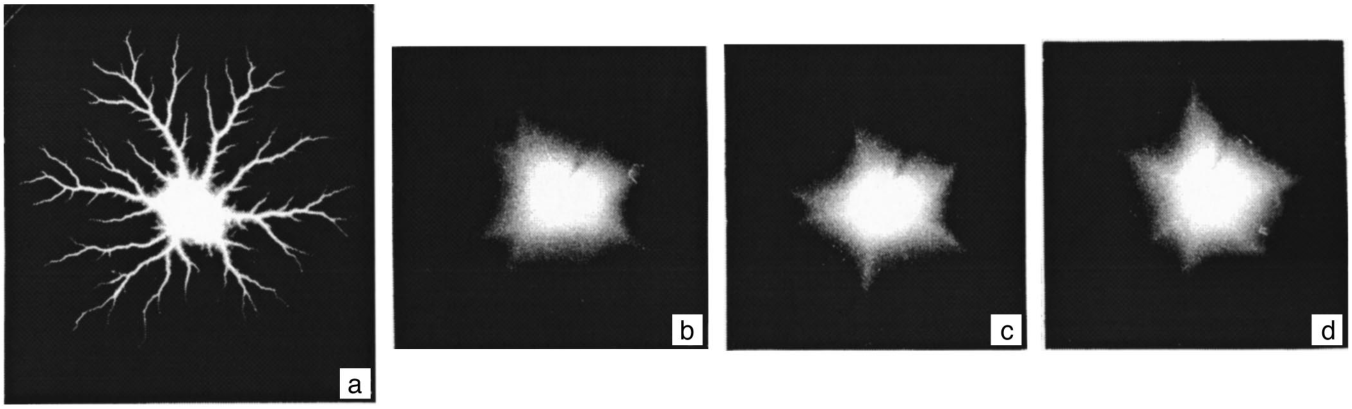

FIG. 3. Discharges with $U_0 = 6$ (a) and 3 kV (b–d).

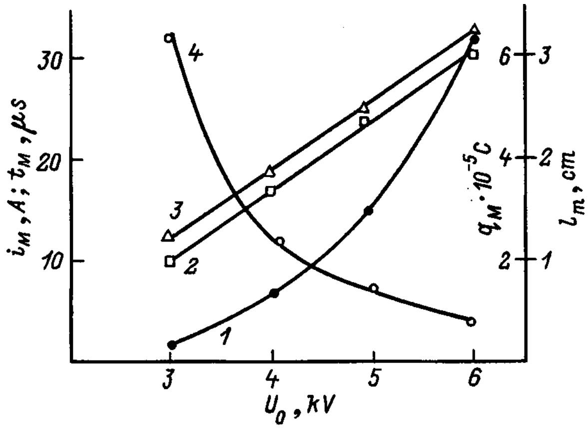

FIG. 4. Maximum values of the current (*1*), channel length (*2*), and charge transferred (*3*) and the time they are attained (*4*) as functions of the initial capacitor voltage.

the amplitude of the current flowing through them. An increase in the area of the channels and the current-conducting layer of water above the cathode leads to a corresponding decrease in the resistance of the discharge circuit (Fig. 2, curves *1* and *3*).

As was established in Ref. 3, the development of a channel is determined by the potential difference between the tip of the channel and the water surface beneath it. The formation of plasma at the tip, as well as along the entire channel, is due mainly to the component of the electric field normal to the water surface, which also determines the flow of current in the water layer. The velocity of the tip of the channel is determined mainly by the longitudinal component of the field, on which the conductivity of the channel also depends. The magnitude of the potential difference between the tip of the channel and the water decreases during development of the channel as a result of the decrease in the capacitor voltage and (mainly) on account of the increase in the voltage drop across the channel itself. This leads to a decrease in the velocity of the tip of the channel and ultimately causes it to stop. At that moment in time the channel length and the current reach their maxima, and the resistance of the discharge circuit drops to its minimum (Fig. 2).

The maximum amplitude of the current in a multichannel discharge was 5–6 times greater than that in a single-channel discharge[3] with the same voltage $U_0 = 6$ kV. This led to more rapid falloff of the capacitor voltage. Accordingly, the potential difference between the tips of the channels and the water surface decreases more rapidly. In addition, since the initial velocities of the tips of the channels should be roughly equal (identical) in the two cases, the length of the channels in the multichannel discharge was smaller than the length of the channel in the single-channel discharge.[3]

After the current in the channel ceased to increase, the power balance was disrupted. The power lost exceeded the power released as a result of the passage of current, primarily at the tips of the channels, with a resulting decay of the plasma in the channels and a decrease in their maximum length. This also triggered the mechanism for the mutual one-to-one correspondence between the current and the channel length, but now in the direction of further reduction of their values.[3]

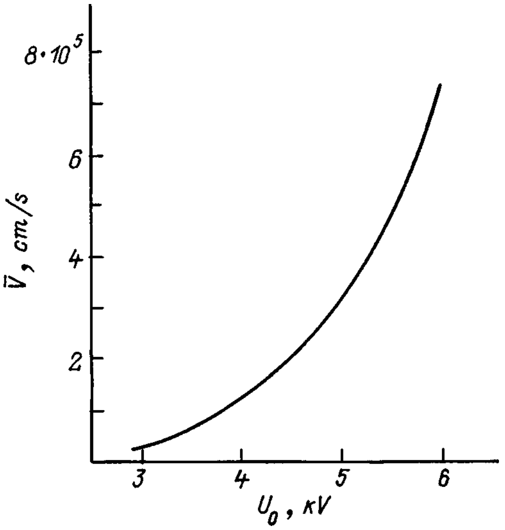

FIG. 5. Mean velocity of channel development as a function of the initial voltage.

The capacitor voltage (Fig. 2, curve *1*) falls off linearly up to the current maximum. As the current decreases, the rate of decrease in the voltage also declines. By the time when the current ceases, the capacitor voltage is equal to approximately one-third of the initial value, and further discharging of the capacitor occurs because of the residual conductivity of the air gap and the water.

The multichannel and symmetric structure of the discharge in the final stages is specified by its corresponding initial structure. This prompts us to examine its inception in greater detail than was done in Refs. 1 and 3. Under the conditions of these studies there was an asymmetric distribution of the field in the discharge gap relative to the anode. Nevertheless, the initial structure of the discharge was also multichannel, and only one channel remained as a result of the competition between the channels and developed under preferential conditions along the axis of the gap. This suggests that the initial multichannel structure of the discharge is related in all cases to the symmetry of the corona discharge, in which the initial channels appeared after the corona discharge touched the water surface. Under the present conditions with a symmetric two-dimensional cathode all the initially appearing channels existed under equal conditions and were capable of developing to completion, thereby facilitating the investigation of this stage. For the same reasons, experiments were carried out at $U_0 = 3$ kV, where the illumination of the central part of the photograph by the corona was minimal.

We, first of all, note that the channels move apart after their appearance (Fig. 3b–3d) and during branching (Fig. 3a). This indicates that the process has an electrostatic character in both cases and that all the channels have a charge of the same sign. Consequently, the entire boundary layer of the plasma, beginning from the corona, has the same positive charge. Therefore, at the moment when the corona cone reaches the negatively charged water surface, a planar double layer forms above it. The field of this layer, which draws electrons from the water surface (the cathode), provides for Ohmic conductivity in the discharge gap from that moment on.

As the current in the circuit increases, the current density in the anode region, with allowance for its diameter, reaches $\simeq 10^4$ A/cm^2 already when $i=0.1$ A. This unavoidably causes reduction of the current in the corona cone. The brightly luminescent channel could be observed visually on the background of the corona. Developing from the anode to the water surface, this instability caused a potential jump upon reaching the double layer and could lead to the emergence of charge-density oscillations and the appearance of a space-charge structure of the Rayleigh–Bénard type. Its maxima could initiate the development of discharge channels. In this case the channel appearing first and each subsequent channel increased the current in the circuit and thereby raised the potential at the center of the structure, promoting the development of subsequent channels. On the other hand, according to Fig. 5 and the data in Ref. 3, the lowering of the capacitor voltage in view of the sharp dependence of the velocity of the chennel tip on it should slow the development of the subsequent channels. These opposing tendencies determined the time interval for the appearance of channels and, consequently, their number, since the time of the appearance of a channel is a random quantity. On the other hand, there is apparently also a spatial constraint on the number of channels, involving the relationship between the perimeter of the corona cone and the channel diameter. Since the current and, therefore, the diameter of the initial channels are smaller at a smaller value of U_0, a greater number could be accommodated on the corona spot when $U_0=3$. The comparatively larger probability of the appearance of four initial channels in all cases (Figs. 3a and 3c) is associated with the character of their electrostatic interaction. Under conditions of an axisymmetric field and a uniform water surface, the second channel more likely appears diametrically opposite the first, and then the third and fourth channels appear along the other diagonal. If two channels azimuthally separated by an angle smaller than 180° were to appear synchronously at first, then a third channel would most probably appear in addition to them to form a stable electrostatic structure (Fig. 3b). Nevertheless, random irregularities on the water surface influence the number of channels and the initial structure.

The integral photographs do not allow us to state that there were actually five or even six initial channels, since this effect could be the result of rapid division of the first channels. Nevertheless, the photographs make it possible to postulate the sequence of the appearance of the channels on the basis of their maximum length and the magnitude of the angle of the sector occupied by channel branches with consideration of the mutual one-to-one correspondence between the current and channel length. Before the beginning of branching, the channels that have appeared develop independently. However, already after the first bifurcation act, they begin to compete for the area of the water surface through which the current flows. The first channel and the channels appearing with a short delay will ultimately develop in a larger sector and have a larger current amplitude. Ultimately, more charge will be transferred through them, and they will attain greater lengths. However, this will occur only under conditions of a symmetric cathode. The wedged shape of the water layer and the consequent asymmetry of its resistance and the current in the channels lead to asymmetry of their structure (Fig. 3a).

As can be seen from Fig. 4, the averaged maximum channel length l_M and the amount of charge transferred in the circuit up to the current maximum q_M exhibit the same dependence on initial voltage. This implies a proportionality between l_M and q_M and fits the conclusion in Ref. 3 that the product of the storage capacitance and the initial potential difference between the tip of the channel and the water surface beneath it, i.e., the charge transferred, is an invariant of channel development. Of course, because of the spread of times for the appearance of channels, the values of l_M and q_M for each channel differ.

Everything we have said regarding the structure of a multichannel discharge, as well as the detailed similarity between the oscillograms of the current and the capacitor voltage for it and the corresponding oscillograms of a single-channel discharge,[3] allow us to assert that the discharge circuit becomes a nonlinear $R(t)C$ circuit from the onset of the development of channels over the water surface in the present case, too. A mutual one-to-one correspondence between the channel length and the current is established during discharge, and it has a self-consistent character as a whole. This allows us to examine the relationship between the energy characteristics and the structure of the leader channels over the surface of a dielectric having some conductivity on the basis of a single theory.

[1] V. P. Belosheev, Zh. Tekh. Fiz. **66**(8), 50 (1996) [Tech. Phys. **41**, 773 (1996)].

[2] S. I. Andreev, E. A. Zobov, and A. I. Sidorov, Prikl. Mekh. Tekh. Fiz. No. 3, 38 (1978).

[3] V. P. Belosheev, Zh. Tekh. Fiz. **68**(7), 44 (1998) [Tech. Phys. **43**, 783 (1998)].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Inhibition of the growth of excrescences in mixed electronic–ionic conductors

M. A. Korzhuev

A. A. Baĭkov Institute of Metallurgy, Russian Academy of Sciences, 117911 Moscow, Russia

(Submitted June 16, 1997)
Zh. Tekh. Fiz. **68**, 67–71 (November 1998)

The thermodynamic aspects of the diagnosis, prevention, and inhibition of the growth of excrescences in mixed electronic–ionic conductors are considered in the case of the representative material copper selenide. © *1998 American Institute of Physics.* [S1063-7842(98)01311-7]

The processes leading to the ''spontaneous'' growth of new formations or ''excrescences'' (a term which we take to include ''moustaches,'' whiskers, ribbons, nodules or ''hillocks,'' pores, cavities, etc.) are currently a focus of research attention. This subject is of interest because the growth of excrescences significantly shortens the service life of materials and instruments in some cases. For example, the formation of pores and cavities leads to the swelling and failure of some structural and functional alloys, and the growth of ''moustaches'' on the surfaces of cables and microcircuits leads to the short-circuiting and malfunctioning of instruments.[1]

The growth of excrescences is stimulated by concentration, temperature, and pressure differentials and by electromagnetic fields, ionizing radiation, and plastic deformation of the samples.[1,2] All other conditions being equal, the growth rate of excrescences is determined by the diffusive mobility of the atoms and ions. The latter reaches a maximum in superionic conductors (solid electrolytes), especially in the presence of mixed electronic–ionic conductivity [in superionic semiconductors of the I–VI group ($A^{I}B^{VI}$, where A = Cu, Ag, Au; B = S, Se, Te) the diffusion coefficients D of the cations are as high as 10^{-1} cm^2/s].[3] Superionic conductors with mixed electronic–ionic conductivity ($Cu_{2-x}Se$ and $Ag_{2-x}Se$), in particular, have previously been found to exhibit anomalously rapid growth of excrescences, based on both the ''molten'' metallic sublattice [copper (silver) ''moustaches,'' ribbons, and drops; the growth rate v was as high as 0.1 mm/h at $T > T_C$, where T_C is the superionic phase transition temperature] and on the entire crystal as a whole (pores, hillocks) (v up to 0.001 mm/h at $T > T_t \sim (0.5-0.8)T_m$, where T_m and T_t are the melting point and the Tammann temperature).[4–9]

Questions concerning the diagnosis, prevention, and inhibition of excrescences growing on surfaces have not previously been specially investigated. The purpose of the present work was to develop methods for the diagnosis, prevention, and inhibition of the growth of excrescences, i.e., copper whiskers and drops on the representative material $Cu_{2-x}Se$[$T_C = 291-430$ K, $T_m = 1380$ K, $x = 0.01-0.3$, D_{Cu} up to 10^{-2} cm^2/s (Refs. 3 and 4)].

EXPERIMENT

Polycrystalline samples of $Cu_{2-x}Se$ with x ranging from -0.03 to $+0.21$, which lie, according to their composition, either near the boundary of the homogeneity region of the copper-saturated compound ($x_0 \sim 0.005$ at $T = 300$ K) or far from it, as well as samples doped to $1-3$ at. % Zn, In, Sb, and In+Sb (Table I), were obtained by ampul synthesis ($t = 1$ h) at 1400 K, and then the alloys were air-cooled to room temperature over the course of $t \sim 40$ min. Molten samples of cylindrical shape ($D = 6$ mm, $h = 16$ mm) with a crystalline grain size equal to $100-300$ μm were used. The samples were annealed in argon at $T = 1073$ K $\sim 0.8T_m$ in a temperature gradient $\nabla T \sim 100$ K/cm over the course of $t = 240$ h. We measured the differential thermopower $\alpha^{300\,K}$ of the samples

$$E = -(\mu_M^0 - \mu_M)/Ze = \frac{\Delta G}{ZF} \tag{2}$$

and the emf of electrochemical cells of the type

$$\mathrm{Cu/aq\ \ CuSO_4/Cu_{2-x}Se}, \tag{3}$$

where μ_{Cu} and μ_{Cu}^0 are the chemical potentials of copper atoms in $Cu_{2-x}Se$ and in metallic copper, ΔG is the corresponding difference between the Gibbs free energies, Ze is the charge of a metal ion in the electrolyte, e is an elementary charge, and F is Faraday's number.[10]

The composition x of the $Cu_{2-x}Se$ matrices were found from the values of $\alpha^{300\,K}$ and E (Ref. 10). The number of phases present in the samples was determined by metallography, and the relative porosity $\Pi = ((d/d_0) - 1)$ of each material was determined by comparing the theoretical (d_0) and experimental (d) densities, the latter being found by hydrostatic suspension of the samples (Table I).

The composition of the alloy matrix was varied in the range $x = 0.005-0.2$ by the extraction and injection of mobile copper. Mobile copper was extracted from the samples by 10% nitric acid [the chemical reaction is $Cu_{2-x}Se + HNO_3 \rightarrow Cu_{2-(x+\Delta x)}Se + Cu(NO_3)_2 + NO(NO_2)\uparrow + H_2O$], and it was injected by self-doping in short-circuited cells of the type (3) with a liquid electrolyte (aq $CuSO_4$).[10]

EXPERIMENTAL RESULTS AND DISCUSSION

The characteristics of the alloys investigated before and after annealing are listed in Table I, and the samples on which excrescences appeared after annealing are marked (+ for copper drops and/or ++ for whiskers). One of the samples, on which both of these types of excrescences grew

TABLE I. Characteristics of the alloys investigated at room temperature before/after annealing ($T = 1073$ K).

Alloys and excrescences	Number of phases	Thermopower α, μV/K	emf E, mV	Porosity, Π
$Cu_{1.99}Se$	1	111/108	79/80	0.98/0.98
$Cu_{2.03}Se^{+,++}$	2*	334/189	32/45	0.97/0.96
$Cu_{1.95}Se$	1	54/52	101/98	0.99/0.99
$Cu_{1.79}Se$	1	9/10	121/119	0.98/0.98
$Cu_{0.636}In_{0.02}Se^{++}_{0.344}$	3*	256/155	27/40	0.96/0.94
$Cu_{0.641}Sb_{0.003}Se_{0.342}$	2	64/63	101/100	0.97/0.96
$Cu_{0.641}Se_{0.342}In_{0.01}Sb^{++}_{0.003}$	3*	290/167	36/61	0.96/0.95
$(Cu_{1.99}Se)_{0.98}(Zn_2Se)^{+,++}_{0.02}$	2*	220/180	28/47	0.95/0.95

* — Copper is present as one of the phases; $^+$ — copper drops; $^{++}$ — whiskers.

simultaneously, is shown in Fig. 1. It can be seen from Table I that the alloys with excrescences consisted of at least two phases prior to annealing and contained metallic copper in the form of a separate phase.

Figures 2a–2c show the relative positions of the levels of the chemical potentials μ_1 of the metal (M) in the metallic phase and μ_2 in the compound (X) in a two-phase M–X system (M denotes Cu, and X denotes $Cu_{2-x}Se$) at a fixed temperature. According to Fig. 2, when there is mechanical contact between the materials, depending on the relative positions of the chemical-potential levels, the copper atoms diffuse from the metallic copper into the $Cu_{2-x}Se$ sample ($\mu_1 > \mu_2$) (a) or in the reverse direction ($\mu_1 < \mu_2$) (c) until a state of equilibrium is established in the system ($\mu_1 = \mu_2$) (b). If the materials in contact are separated by an electron filter, i.e., a copper-conducting electrolyte (e), which passes Cu^+ ions and holds back electronic charge carriers, the emf E (2) appears on the open contacts of the electrochemical cell (3), and its sign indicates the relative positions of the chemical-potential levels in the phases in contact [$E > 0$ when $\mu_1 > \mu_2$ (Fig. 2a), $E < 0$ when $\mu_1 < \mu_2$ (c), and $E = 0$ when $\mu_1 = \mu_2$ (2d–2f)].

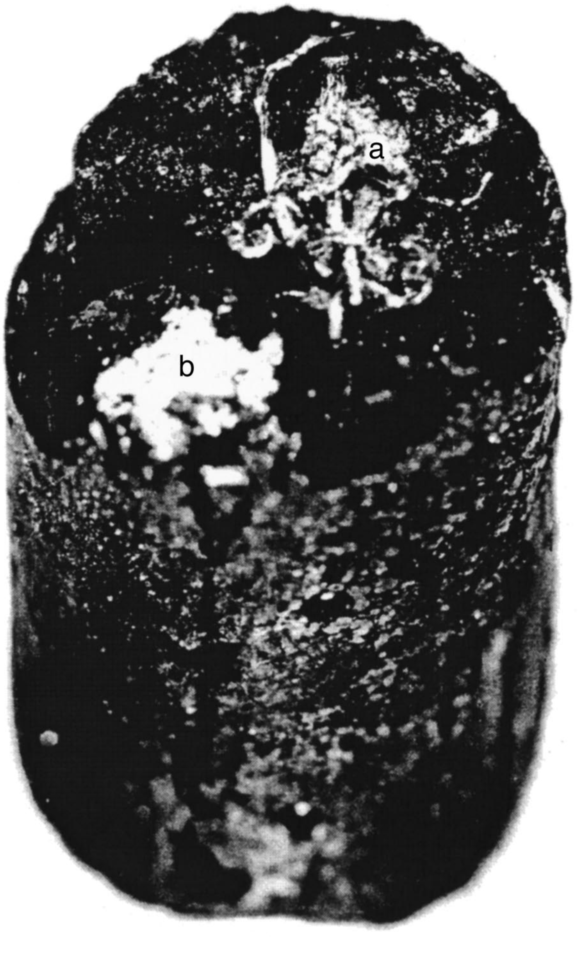

FIG. 1. Sample of $Cu_{2-x}Se$ ($x = -0.03$) with a colony of filamentary copper crystals (a) and copper drops (b) formed as a result of annealing ($t = 240$ h, $T = 1073$ K) on the surface ($10^\times$).

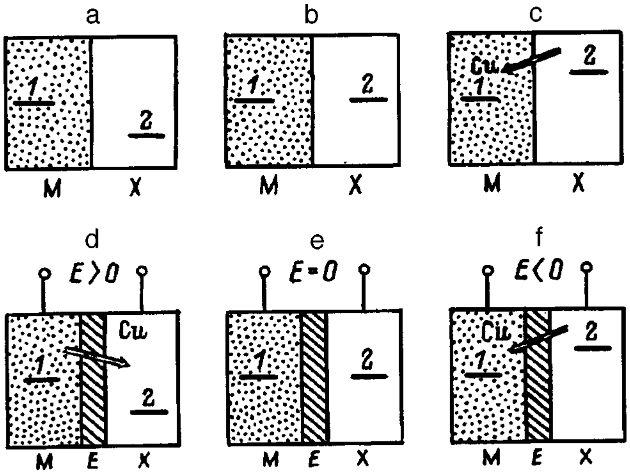

FIG. 2. Possible arrangements of the chemical-potential levels of a metal in the metallic phase and in a compound in M–X and M–E–X systems.

When $E < 0$ (Fig. 2c), the system is supersaturated with copper, and the nucleation and growth of excrescences (M) can occur on the sample (X). When $E > 0$ (a), the excrescences (M) growing on the sample dissolve in the solid solution (X), and when $E = 0$, the excrescences and the sample are in a state of equilibrium.[3] It can be seen from Table I that values in the range $E \gtrsim 0$ are typical of alloys with excrescences both before and after annealing. Hence it follows that: 1) during cooling after synthesis and storage at room temperature, the supersaturated solid solutions based on copper selenide have already attained an equilibrium state by undergoing the decomposition reaction ($Cu_{2-x'}Se \rightarrow Cu_{2-x_0}Se + Cu$) with the formation of "primary excrescences" (internal and surface microdeposits of copper) ($D \sim 1-100$ μm; Refs. 5 and 10); 2) the whiskers and drops are "secondary excrescences," which grow as a result of the migration of copper in the primary excrescences to the boule surface during annealing under the action of external forces (temperature gradients, pressure gradients, etc.) according to the scheme $Cu_I \rightarrow Cu_{2-x}Se \rightarrow Cu_{II}$; and 3) the nucleation centers for the primary excrescences in $Cu_{2-x}Se$ are pores and the sample surface,[5,6] while the nucleation centers for the secondary excrescences (Fig. 1) are the "primary" surface deposits of copper. The role of the external forces reduces here to ensuring the condition $E < 0$ in the regions where copper condenses. When $E \sim 30$ mV, the condition $E < 0$ is achieved in the samples (Table I) if there is a local increase in the

TABLE II. Various diffusion effects observed in $Cu_{2-x}Se$ ($T=300$ K).

Type of effect	Magnitude of effect ($\delta C_{Cu}/\delta Y$)	Reference
Thermodiffusion	6×10^{17}, K^{-1} cm^{-3}	Ref. 11
Piezelectric diffusion:		
a — hydrostatic pressure	-3×10^{16}, MPa/cm^3	Ref. 7
b — uniaxial compression	-7×10^{16}, MPa/cm^3	Ref. 11
Electrodiffusion	2×10^{19}, mm^2/(A·cm^3)	Ref. 12
Gravitational	$\sim10^{14}$, S^2/cm^4	Ref. 13

copper concentration $\Delta C_{Cu}\geqslant10^{18}$ cm^{-3}. Such values of ΔC_{Cu} can be caused by a temperature difference $\Delta T\sim10$ K, a pressure difference $\Delta P\sim0.3-0.6$ GPa,[11] an electric current of density $j\sim0.1$ A/mm^2 (Ref. 12), and acceleration of the sample at a rate $g\sim10^3 g_0$ (Ref. 13) (see Table II). The relative smallness of these values shows that the risk factor for the growth of excrescences on samples of $Cu_{2-x}Se$ is fairly high when $x\sim x_0$.

The increased tendency for the growth of excrescences is associated with the special state of the ''molten'' superionic sublattice and the high diffusive mobility of the respective ions.[4] It is known that mobile copper ions are in a more disordered state in the ''molten'' $Cu_{2-x}Se$ sublattice ($T>T_C$) than in the liquid phase. Therefore, at the same degree of supercooling (supersaturation) the condensation process of mobile copper according to the ''molten'' $Cu_{2-x}Se$ sublattice→crystal scheme corresponds to a greater entropy change than does crystallization from a melt (by a factor of ~3), and the thermodynamic stimulus of the formation of copper nuclei in $Cu_{2-x}Se$ as a result of fluctuations increases accordingly.

On the other hand, the probability of fluctuations in a system which is not too close to the critical point is specified by the Einstein formula

$$W\sim\exp((S''-S)/\mathrm{k}_0), \tag{4}$$

where S and S'' are the entropies of the system in the equilibrium state and in the fluctuation state, and k_0 is Boltzmann's constant.[14,15]

Since the quantity $\alpha'_m=((S''_{Cu}-S_{Cu})/\mathrm{k}_0)$ (here S''_{Cu} and S_{Cu} are the entropies of copper in the ''molten'' $Cu_{2-x}Se$ sublattice and in the copper excrescences) is increased significantly,[16] the probability of the formation of copper nuclei as a result of fluctuations is diminished, and the probability of the resorption of a copper excrescence is increased in comparison to ordinary materials. For just this reason: 1) primary excrescences form in $Cu_{2-x}Se$ exclusively at sites of defects (pores, surfaces);[5,6] this is promoted by the relative porosity of the boules (Table I), and the high diffusive mobility of the copper ions specifies the rapid achievement of critical dimensions by the nuclei of the excrescences;[14] as a whole, the formation of copper nuclei in the samples is facilitated significantly, as is indicated by the virtual absence of an incubation period for decomposition of the corresponding solid solution;[10] 2) the primary excrescences readily dissolve in the solid solutions in $Cu_{2-x}Se$, creating, in particular, a possibility for the rapid growth of secondary excrescences based on the primary excrescences according to the scheme $Cu_I\rightarrow Cu_{2-x}Se\rightarrow Cu_{II}$.

The rapid growth rate of all the excrescences investigated in $Cu_{2-x}Se$ is associated with the high diffusive mobility of copper in the samples.[4] The characteristic times for the diffusion of copper are

$$\tau\sim d^2/(\pi^2 D), \tag{5}$$

where D is the diffusion coefficient of mobile copper and d is the diffusion length.[3] If the distance between microdeposits $d\sim100$ μm and $D^{300\,K(1000\,K)}\sim10^{-6(-3)}$ cm^2/s (Refs. 3 and 17), this formula gives the following estimates of the decomposition time of a supersaturated solid solution of copper in $Cu_{2-x}Se$: ~10 h and ~1 min at 300 and 1000 K, respectively.[18]

For secondary excrescences $d\sim1$ cm, whence we obtain the estimate for τ of several months and several hours at $T=300$ and 1000 K, respectively. The results obtained in the present study permit formulation of the main principles for the diagnosis, prevention, and inhibition of the growth of excrescences on samples of $Cu_{2-x}Se$.

Diagnosis of excrescences. The condition $E<0$ can serve as a criterion for the possible growth of primary excrescences (copper microdeposits) on samples of $Cu_{2-x}Se$ (Fig. 2f). The states of samples with $E<0$ are unstable and exist for a relatively short time (~1 min to several hours), after which the solid solution decomposes as $E\rightarrow0$ (Fig. 2e).[10] The condition $E\sim0$ and the presence of metallic copper in the alloy in the form of a separate phase can serve as a criterion for the possible growth of secondary excrescences (copper ''moustaches'' and drops) on the samples of $Cu_{2-x}Se$ (Table I).

Treatment of excrescences. Copper excrescences on a $Cu_{2-x}Se$ surface can be removed by nitric acid [in the chemical reaction $Cu+HNO_3\rightarrow+Cu(NO_3)_2+NO(NO_2)\uparrow+H_2O$] ($t\sim1$ h) or dissolve in the solid solution (in the chemical reaction $Cu+Cu_{2-x}Se\rightarrow Cu_{2-x'}Se$, $x>x'$).[10] To realize the latter possibility, the level of the chemical potential of copper in $Cu_{2-x}Se$ must be lowered until $\mu_1>\mu_2$ (Fig. 2a).

The level of μ_2 was lowered in the present study by extracting mobile copper from the sample with nitric acid ($x\rightarrow0.20$).[10] Figure 3 shows how the thermopower α (curve *1*), the mass m/m_0 (curve *2*), and the emf E (curve *3*) of a sample of $Cu_{2-x}Se$ ($x=-0.03$) vary in this process ($b\rightarrow c$). Then the sample with excrescences was subjected to an homogenizing anneal in argon ($T=1000$ K, $t=100$ h). During the anneal, however, the excrescences were largely removed from the sample.

In order to stabilize the mass-transfer process, the sample was placed in the copper-conducting electrolyte aq $CuSO_4$, where it was held for 24 h ($T=300$ K) (Fig. 3d). This produced a short-circuited galvanic cell of the type (3), under whose emf E the copper in the excrescences dissolved in the electrolyte and was then injected by the electric field of the cell E into the solid solution through the free surface of the sample (Fig. 3d).[10] Complete resorption of the excres-

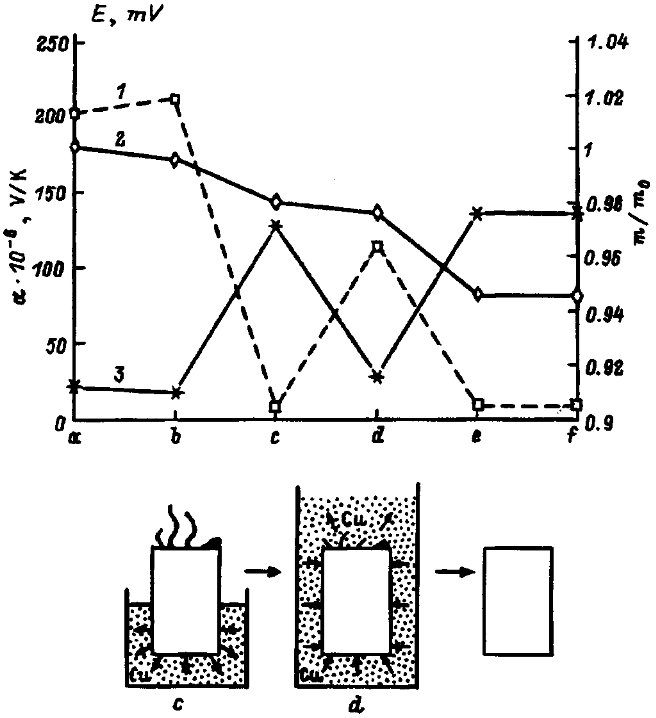

FIG. 3. The thermopower α (*1*), the relative mass change m/m_0 (*2*), and the emf E (*3*) of a $Cu_{2-x}Se/aqCuSO_4/Cu$ electrochemical cell with a sample of copper selenide ($x=-0.03$) in various stages of its processing (*a–f*): *a* — synthesis; *b* — growth of copper filaments; *c,e* — extraction of copper by nitric acid; *d,f* — saturation with copper in short-circuited electrochemical cells (*3*).

cences required 5–16 h, during which the total mass of the sample remained essentially unchanged (Fig. 3, $c \rightarrow d$, curve *2*).

Prevention of the growth of excrescences. It can be seen from Table I that copper excrescences do not grow on samples of $Cu_{2-x}Se$ with $x \geqslant 0.05$ ($\mu_1 \gg \mu_2$; Fig. 3a). Therefore, the prevention of the growth of both primary and secondary excrescences can include shifting the composition of the solid solution into the safe range (in practice to $x > 0.02 - 0.03$) by extracting copper from the sample. It is also possible to introduce dopants which form chemical compounds with copper and thus eliminate the possibility of the formation of copper nuclei on the boules (antimony-doped alloys, Table I). The influence of external forces on the samples should be eliminated in all cases (Table I). The critical values of external forces capable of causing the growth of excrescences on a specific sample can be estimated from the condition

$$Y_c \sim (x - x_0)/k, \tag{6}$$

where x and x_0 are the current alloy composition and the composition corresponding to the boundary of the homogeneity region, and $k = (\delta C_{Cu}/\delta Y)/(1.99\times 10^{22})$ is a dimensionless coefficient which describes the corresponding diffusion effect (Table II).

It can be seen from (6) that under the action of strong external forces ($Y \gg Y_C$) excrescences can appear not only on samples with $x \sim x_0$ ($E \sim 0$), but also on samples with $x > x_0$ ($E > 0$), as was previously observed[9] in the plastic deformation of samples of $Cu_{1.99}Se$.

CONCLUSION

In summary, we have developed methods for the diagnosis, prevention, and treatment of excrescences, i.e., copper whiskers and drops formed during annealing on boules of the nonstoichiometric superionic conductor copper selenide $Cu_{2-x}Se$ and solid solutions based on it (Table I). We have shown that growing excrescences can be classified as primary and secondary formations, which have different diagnostic criteria ($E < 0$ and $E \sim 0$) (Table I). The criteria indicated can be used to completely eliminate the growth of copper excrescences on samples of $Cu_{2-x}Se$ by shifting the composition of the solid solution into the safe range. At the same time, copper excrescences growing on samples of $Cu_{2-x}Se$ can be eradicated by stimulating reverse mass transfer with the sample (Fig. 3).

The results obtained in this work can also be applied to other superionic conductors. They are also valid for an extensive list of solids, on which excrescences grow under a mechanism of diffusive transport of a substance.

We thank V. F. Bankina and N. N. Filipovich for synthesizing the samples investigated in this work.

[1] G. V. Berezhkova, *Filamentary Crystals* [in Russian], Nauka, Moscow (1969), 158 pp.

[2] Ya. E. Geguzin, *The Diffusion Zone* [in Russian], Nauka, Moscow (1979), 344 pp.

[3] V. N. Chebotin, *Chemical Diffusion in Solids* [in Russian], Nauka, Moscow (1989), 208 pp.

[4] M. A. Korzhuev, JETP Lett. **64**, 105 (1996).

[5] M. A. Korzhuev, N. Kh. Abrikosov, and V. F. Bankina, Pis'ma Zh. Tekh. Fiz. **10**, 1486 (1984) [Sov. Tech. Phys. Lett. **10**, 627 (1984)].

[6] M. A. Korzhuev, V. F. Bankina, N. Kh. Abrikosov *et al.*, Zh. Tekh. Fiz. **56**, 797 (1986) [Sov. Phys. Tech. Phys. **31**, 484 (1986)].

[7] M. A. Korzhuev, N. Kh. Abrikosov, and I. V. Kuznetsov, Pis'ma Zh. Tekh. Fiz. **13**, 9 (1987) [Sov. Tech. Phys. Lett. **13**, 4 (1987)].

[8] M. A. Korzhuev, B. A. Efimova, E. A. Obraztsova, and O. P. Fedorova, Phys. Status Solidi A **124**, 115 (1991).

[9] M. A. Korzhuev, Fiz. Khim. Obrab. Mater. No. 5, 153 (1993).

[10] M. A. Korzhuev, Fiz. Khim. Obrab. Mater. No. 1, 92 (1989); Fiz. Khim. Obrab. Mater. No. 2, 137 (1989); Fiz. Khim. Obrab. Mater. No. 5, 145 (1990).

[11] M. A. Korzhuev and A. V. Laptev, Fiz. Tverd. Tela (Leningrad) **29**, 2646 (1987) [Sov. Phys. Solid State **29**, 1524 (1987)].

[12] M. A. Korzhuev and A. V. Laptev, Zh. Tekh. Fiz. **59**(4), 62 (1989) [Sov. Phys. Tech. Phys. **34**, 419 (1989)].

[13] M. A. Korzhuev, Fiz. Tverd. Tela (Leningrad) **30**, 2387 (1988) [Sov. Phys. Solid State **30**, 1376 (1988)].

[14] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 2nd ed. (Pergamon Press, Oxford, 1969; Mir, Moscow, 1964).

[15] A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions* (Pergamon Press, Oxford, 1979; Nauka, Moscow, 1975, 256 pp.).

[16] M. A. Korzhuev, V. V. Baranchikov, N. Kh. Abrikosov, and V. F. Bankina, Fiz. Tverd. Tela (Leningrad) **26**, 2209 (1984) [Sov. Phys. Solid State **26**, 1341 (1984)].

[17] R. A. Yakshibaev, V. N. Konev, and M. Kh. Balapanov, Fiz. Tverd. Tela (Leningrad) **26**, 3641 (1984) [Sov. Phys. Solid State **26**, 2189 (1984)].

[18] M. A. Korzhuev, Fiz. Khim. Obrab. Mater. No. 6, 138 (1990).

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Calculation of the permeability of polycrystalline ferrites

L. N. Kotov and K. Yu. Bazhukov

Syktyvkar State University, 167001 Syktyvkar, Russia

(Submitted July 8, 1997)
Zh. Tekh. Fiz. **68**, 72–75 (November 1998)

[S1063-7842(98)01411-1]

One of the main properties characterizing the behavior of magnetic materials in variable magnetic fields is the magnetic permeability μ, which is a complex quantity: $\mu=\mu' - i\mu''$. The frequency dependence of the permeability is influenced mainly by two processes: motion of the domain boundaries, and rotation of the magnetization vector. There are numerous models that are used to describe and explain the behavior of the permeability as the frequency is varied.[1–3] However, the calculations based on these models generally provide adequate descriptions of the frequency dependence of the permeability only in a narrow frequency range.[3] This may be because most of the models considered do not take into account the rotation of the magnetization vector. At high frequencies (for example, frequencies above 10^8 Hz for yttrium iron garnet)[1] the influence of the rotation of the magnetization vector becomes comparable to the influence of the motion of the domain boundaries and even exceeds it, and at low frequencies the maximum contribution of the rotation of the magnetization vector is specified by the quantity $\chi\approx M_S/H_A$ (where M_S is the saturation magnetization and H_A is the anisotropy field) and can amount to only 20% of the contribution of the motion of the domain boundaries.[3] In other models only the rotation of the magnetization vector is considered, and consequently these models describe the experimental data at high frequencies.[1] This paper proposes a model for calculating the permeability with consideration of the contributions of both the motion of the domain boundaries and the rotation of the magnetization vector over a broad frequency range for polycrystalline ferrites. The permeability is calculated as the sum of two contributions: $\mu=\mu_{\rm dom}+\mu_{\rm rot}$, where $\mu_{\rm dom}$ is the permeability due to the motion of the domain boundaries and $\mu_{\rm rot}$ is the permeability due to the rotation of the magnetization vector. The calculations are performed under the assumption that the external magnetic field H_0 is equal to zero and the anisotropy field H_A is greater than $4\pi M_S$, as is typical of the independent-grain model.[1] The applicability of the model is demonstrated in the example of yttrium iron garnet with an aluminum impurity, for which the independent-grain condition holds.

Each domain boundary is characterized by its own resonance frequency f_0. With consideration of the independent-grain model, we assume that the spatial orientation of the grain boundaries is random in a polycrystalline medium. Then, with allowance for the eigenfrequency distribution function $\varphi(f_0)$ of the domain boundaries, the mean permeability due to the motion of the domain boundaries can be represented in the form[3]

$$\mu'_{\rm dom}(f)=1+4\pi B \times\int_{f_{0\rm min}}^{f_{0\rm max}} f_0\,\frac{f_0^2-f^2+4Ef^2\alpha_{\rm elas}^2}{(f_0^2-f^2)^2+4\alpha_{\rm elas}^2 f_0^2 f^2}\,\varphi(f_0)\,df_0,$$

$$\mu''_{\rm dom}(f)=8\pi B\alpha f \times\int_{f_{0\rm min}}^{f_{0\rm max}} \frac{f_0^2-Ef_0^2+Ef^2}{(f_0^2-f^2)+4\alpha_{\rm elas}^2 f_0^2 f^2}\,\varphi(f_0)\,df_0, \quad (1)$$

where $f=\omega/2\pi$, $B=C^*M_S f_0^*/4\pi f_u$, $C^*=9.45$ MHz/Oe, $\alpha_{\rm elas}$ is the quasielastic coefficient of the domain boundaries, f_u is the frequency for maximum absorption in the experimental spectrum, f_0^* is the effective resonance frequency, $\varphi(f_0)\,df_0$ specifies the fraction of domain boundaries with a resonance frequency in the range from f_0 to f_0+df_0 [$\varphi(f_0)$ must be normalized], $f_{0\,\rm min}$ is the minimum resonance frequency of the domain boundaries, and $f_{0\,\rm max}$ is the maximum resonance frequency of the domain boundaries.

The form of the eigenfrequency distribution function of the domain boundaries $\varphi(f_0)$ shown in Fig. 1 was taken from Ref. 3. The function $\varphi(f_0)$ should have the form of a Poisson distribution, but to speed up the calculation process, it was represented in the form of a broken line, in the manner of Ref. 3. The normalization constants, the values of $f_{0\,\rm min}$ and $f_{0\,\rm max}$, and the intermediate frequencies on the broken curve f_1, f_2, f_3, f_4, and f_5 were selected experimentally.

The components of the permeability μ due to the rotation of the magnetization vector have the following form:[1]

$$\mu'_{\rm rot}=1+4\pi(\gamma M_S f_0(f_0^2-(1-\alpha^2)f^2)) \times[(f_0^2-(1+\alpha^2)f^2)^2+4\alpha^2 f^2 f_0^2]^{-1},$$

$$\mu''_{\rm rot}=4\pi(\alpha\gamma M_S f(f_0^2+(1+\alpha^2)f^2)) \times[(f_0^2-(1+\alpha^2)f^2)^2+4\alpha^2 f^2 f_0^2]^{-1}, \quad (2)$$

where $\gamma=2\pi\cdot 2.8$ MHz/Oe is the gyromagnetic ratio, M_S is the saturation magnetization, $\alpha=f_r/f_0$ is the dissipation parameter, f_0 is the ferromagnetic resonance frequency, f_r is the relaxation frequency, and f is the frequency of the magnetic field.

The frequencies f_0 and f_r depend on the effective field, $H_{\rm eff}$, acting on the magnetic moment in the respective particle of the polycrystal: $f_0=\gamma/2\pi H_{\rm eff}$, $f_r=f_r(H_{\rm eff})$.[4] It is

1063-7842/98/43(11)/4/$15.00

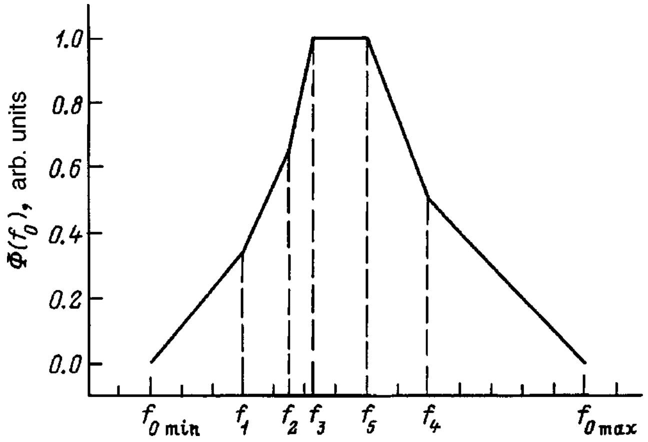

FIG. 1. Differential distribution function of the resonance frequencies of domain boundaries.

assumed that $f_r(H_{\text{eff}})$ is a linear function; therefore, the dissipation parameter α will be regarded as a constant. In solving the problem, α is varied until agreement between the experimental and theoretical magnetic spectra is obtained.

The nonuniformities of the internal field within the grains have different scales. The largest nonuniformities are associated with the shape of the sample. We shall consider domains in the form of parallelepipeds or cylinders. This will be done for the following reasons: a) the simplicity of the formation of domain structures in the crystal without significant distortion of the field near a common boundary between neighboring domains; b) the possibility of considering the variation of the field along one preferred axis, i.e., the Z axis, which makes the most significant contribution to the behavior of the field; c) the fact that the dependence of the magnetic field strength along the field axis in the case of magnetic saturation is well known, and the variation of the transverse field components can be neglected.

For a magnetized sample (a domain in our case) of length L and width D the demagnetizing field can be represented in the form[5]

$$\frac{H_R(0.\xi)}{2\pi M_S} = -2 + \frac{1-\xi}{[k^2+(1-\xi)^2]^{1/2}} + \frac{1+\xi}{[k^2+(1+\xi)^2]^{1/2}}, \tag{3}$$

where $\xi = 2z/L$, $k = D/L$, and the Z axis is directed along the axis of the parallelepiped.

The distribution of the internal magnetic field in this case has the form $H(z) = H_A + H_R(z)$, where H_R is the demagnetizing field and H_A is the anisotropy field.

We have described the internal magnetic field that is characteristic of a single grain with a single domain. When the external field $H_0 = 0$, a ferrite polycrystal has a multidomain structure. If we assume that there are only 180° domain boundaries, then in each individual grain the magnetization vectors **M** in neighboring domains are antiparallel; therefore, the number of domains in a grain must be even, due to the presence of two types of 180° domains. We shall assume that each grain in a polycrystal consists of two 180° domains and closing domains. The contribution of the closing domains to the susceptibility is negligible and will not be considered further. An increase in the number of domain pairs in a polycrystal grain leads only to a decrease in the domain width D and does not influence the final result in our model, altering only the domain-width distribution function. We assume that the domain width and length satisfy the condition $D/L \ll 1$ (i.e., $D \leqslant 0.01L$). We performed a numerical experiment, which revealed that a decrease in the ratio D/L, beginning at 0.01, has practically no influence on the final result (the values vary within less than 1%). For this reason, we shall not examine the width distribution. The spread of domain lengths (which is simultaneously the spread of grain lengths) will be taken into account by a distribution function. The lengths L of an individually taken grain and a domain obey a certain distribution $f(L)$, which must satisfy the following conditions: a) the function must be normalized, i.e., $\int_0^\infty f(L)\,dL = 1$, and b) the following boundary conditions must be satisfied: $f(L) = 0$ at $L = 0$ and $f(L) \to 0$ at $L \to \infty$.

A Poisson law was selected for the distribution function. When the conditions (a) and (b) are taken into account, $f(L)$ has the following form: $f(L) = (L/\langle L\rangle^2)\exp(-L^2/2\langle L\rangle^2)$, where $\langle L\rangle$ is the mean length of the domains. After finding the susceptibility of each domain, averaging is performed over the polycrystal. The chaotic orientation of the magnetization vectors **M** of the grains relative to the axis along which the variable magnetic field propagates (the Z axis) must be taken into account during the averaging. The value of the projection of the internal (effective) field H_{eff} along the Z axis for randomly oriented grains can be represented in the form

$$H_{\text{eff}} = |\mathbf{H}_{\text{eff}}| \cos(2\pi\psi), \tag{4}$$

where ψ, i.e., the angle between the magnetization vector **M** and the Z axis, obeys a Gaussian distribution.

The choice of a Gaussian distribution is most applicable to the independent-grain model ($H_A > 4\pi M_S$), which we are considering in this paper. Various additional fields, which can be associated with various inhomogeneities within the crystallites, can also appear in the grains. Consequently, there is always some constant mean field at the center of a grain. We shall call it the anisotropy field, since it is this field that makes the main contribution. A typical value of the anisotropy field — that of yttrium iron garnet at room temperature, for example — is ~ 80 Oe.[6]

The internal magnetic field is nonuniform in most cases, and this must be taken into account. In formulas (2) there are quantities which depend on the internal field. In order to find the permeability of a substance in a sample in the shape of a parallelepiped with a nonuniform internal field along the Z axis, averaging must be performed over the volume of one domain. For the integration we divide the parallelepiped into thin layers of thickness dZ (within which we can regard H_{eff} as a uniform field):

$$\frac{d\mu'}{\mu'} = \frac{dV}{V}, \quad d\mu' = \mu' \frac{dV}{V}, \quad \langle \mu' \rangle = \int_V \mu' \frac{dV}{V}, \tag{5}$$

where $d\mu'$ and dV are the magnetic permeability and the volume of a thin layer of the parallelepiped, and μ' and V

are the permeability when there is a uniform field throughout the entire parallelepiped and the volume of the crystal.

From (2) and (5) we obtain

$$\langle \mu'_{\mathrm{rot}} \rangle = 1 + 4\pi \times \int_V \frac{\gamma^2 M_S H_{\mathrm{eff}}/2\pi((2\gamma/2\pi)^2 H_{\mathrm{eff}}^2 - (1-\alpha^2) f^2)}{((\gamma/2\pi)^2 H_{\mathrm{eff}}^2 - (1+\alpha^2) f^2)^2 + 4\alpha^2 (\gamma/2\pi)^2 H_{\mathrm{eff}}^2 f^2} \frac{dV}{V}, \tag{6}$$

where $V = D^2 L$, $dV = D^2\, dL$, and $dL = dZ$.

Since the cross section of the parallelepiped is constant, we can go from an integral over volume to an integral over the length L, and then formula (6) ultimately takes the form

$$\langle \mu'_{\mathrm{rot}} \rangle = 1 + 4\pi \times \int_{-L/2}^{L/2} \frac{\gamma^2 M_S H_{\mathrm{eff}}/2\pi((\gamma/2\pi)^2 H_{\mathrm{eff}}^2 - (1-\alpha^2) f^2)}{((\gamma/2\pi)^2 H_{\mathrm{eff}}^2 - (1+\alpha^2) f^2)^2 + 4\alpha^2 (\gamma/2\pi)^2 H_{\mathrm{eff}}^2 f^2} \frac{dZ}{L},$$

$$\langle \mu''_{\mathrm{rot}} \rangle = 4\pi \times \int_{-L/2}^{L/2} \frac{\alpha \gamma M_S f((\gamma/2\pi)^2 H_{\mathrm{eff}}^2 + (1+\alpha^2) f^2)}{((\gamma/2\pi)^2 H_{\mathrm{eff}}^2 - (1+\alpha^2) f^2)^2 + 4\alpha^2 (\gamma/2\pi)^2 H_{\mathrm{eff}}^2 f^2} \frac{dZ}{L}. \tag{7}$$

The integrals (7) cannot be solved analytically. Therefore, the problem of calculating the permeability was reduced to numerical integration. In the next step the mean permeability for an entire polycrystal is found with allowance for the fact that the domain length L varies according to a distribution function.

Calculations were performed for samples of yttrium iron garnet containing aluminum impurities with the composition $Y_2O_3(5-X)Fe_2O_3XAl_2O_3$ ($X = 0.7$ and 1.5). The anisotropy field ($H_A \approx 80$ Oe), the values of the saturation magnetization ($M_S = 61$ and 13 G), and experimental frequency dependences of the permeability over a broad frequency range are known for them.[6] In addition, the loss peaks caused by the motion of the domain boundaries and the rotation of the magnetization vector are observed in different frequency ranges for these samples. This is undoubtedly important for refining the present model. Variation of the grain-length distribution function showed that the mean grain length $\langle L \rangle = 2.5\ \mu$m is optimal for these samples. In fact, the dimensions of the particles in polycrystalline ferrites have the mean value just indicated.[6] In our model we adopted the condition $D/L \ll 1$, which also holds for yttrium iron garnet with an aluminum impurity, because the experimental data used were obtained on a sample in the form of a torus.[7] The mean magnetic field existing within the grains of the polycrystal was estimated. According to the results obtained, the mean field (the anisotropy field, the demagnetizing field, and the magnetostriction field) is identical for the yttrium ion garnets of both compositions and amounts to 157 Oe, i.e., the mean field at the center of the polycrystal grains generally exceeds the anisotropy field ($H_A \approx 80$ Oe). The minimum and maximum resonance frequencies of the domain-boundary motion for the sample with $X = 0.7$ are $f_{0\min} = 13$ MHz and $f_{0\max} = 51$ MHz with the intermediate frequencies (in MHz) along the broken curve $f_1 = 13.2$ MHz, $f_2 = 13.5$ MHz, $f_3 = 13.8$ MHz, $f_4 = 15.6$ MHz, and $f_5 = 18$ MHz and the absorption frequency $f_u = 10^7$ MHz. The values for the sample with $X = 1.5$ are $f_{0\min} = 20$ MHz, $f_{0\max} = 58$ MHz, $f_1 = 20.2$ MHz, $f_2 = 20.5$ MHz, $f_3 = 20.8$ MHz, $f_4 = 22.6$ MHz, $f_5 = 25$ MHz, and $f_u = 2.3 \times 10^7$ MHz. The parameter $E = 1$ and the quasi-elastic coefficient $\alpha_{\mathrm{elas}} = 0.56$ (Oe·G)/cm were used for both samples. It became clear during further numerical experiments that the dissipation coefficient $\alpha = 0.8$ and the spin relaxation frequency $\omega_r = 2\pi f_r = 0.8\omega_0$ for the yttrium iron garnet with $X = 0.7$, while $\alpha = 0.57$ and $\omega_r = 2\pi f_r = 0.57\omega_0$, respectively, for the sample with $X = 1.5$, in good agreement with the data obtained in Ref. 8.

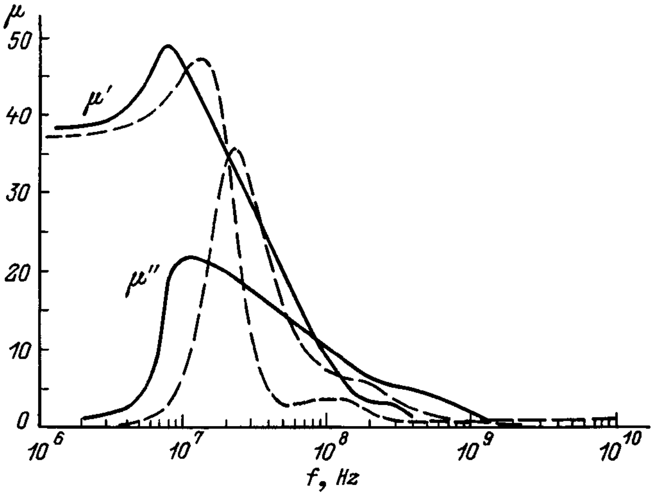

FIG. 2. Frequency dependence of the real and imaginary parts of the magnetic susceptibility for the yttrium ion garnet with $X = 0.7$.

Figure 2 shows plots of the frequency dependence of the real and imaginary parts of the magnetic permeability of the yttrium ion garnet $Y_2O_3(5-X)Fe_2O_3XAl_2O_3$ with $X = 0.7$. The solid line in Fig. 2 corresponds to the experimental data, and the dashed line corresponds to the theoretical values obtained using the independent-grain model. It is seen that the forms of the theoretical and experimental plots coincide poorly with one another, although the experimental and theoretical values of the real part of the permeability at different frequencies are close (the differences amount to no more

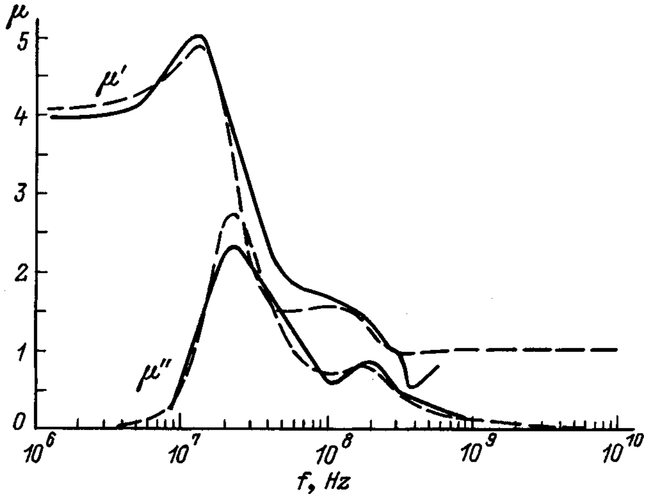

FIG. 3. Same as in Fig. 2 for the sample with $X = 1.5$.

than 20%). This is attributed to the fact that the grain-independence condition $H_A > 4\pi M_S$ does not hold for this ferrite. As can be seen from Fig. 2, the interaction of the magnetic subsystems of the grains in this ferrite leads to alteration of the resonance frequencies and the recombination frequencies. In order to allow for the influence of the grains on one another, another distribution function, which takes into account the magnetic interaction between domains in neighboring grains, must be introduced into the model.

The plots shown in Fig. 3 were obtained for the yttrium ion garnet $Y_2O_3(5-X)Fe_2O_3XAl_2O_3$ with $X=1.5$ (the solid line shows the experimental data, and the dashed line is the theoretical result). It can be seen from Fig. 3 that the frequency dependences for this ferrite agree very well with the experimental data. Thus, for polycrystals in which the condition $H_A > 4\pi M_S$ holds, the model of a polycrystal with independent grains, in which both the motion of the domain boundaries and the rotation of the magnetization vector are taken into account, permits obtaining the frequency dependences of the magnetic permeability to within good accuracy. This is undoubtedly important for using this model to calculate the magnetic permeability of polycrystalline ferrite media with specified frequency properties.

[1] A. S. Gurevich, *Magnetic Resonance in Ferrites and Antiferromagnets* [in Russian], Nauka, Moscow (1973), 592 pp.
[2] S. Krupicka, *Physik der Ferrite und der Verwandten Magnetischen Oxide* (Academia, Prague, 1973; Mir, Moscow 1976, 504 pp.).
[3] G. Zh. Rankis, *Dynamics of the Magnetization of Polycrystalline Ferrites* [in Russian], Zinatne, Riga (1981), 384 pp.
[4] B. A. Goldin, L. N. Kotov, L. K. Zarembo, and S. N. Karpachev, *Spin-Phonon Interactions in Crystals (Ferrites)* [in Russian], Nauka, Leningrad (1991), 150 pp.
[5] G. A. Smolenskiĭ, V. V. Lemanov, G. M. Nedelin *et al*., *Physics of Magnetic Dielectrics* [in Russian], Nauka, Leningrad (1974), 334 pp.
[6] B. M. Lebed' and S. G. Abarenkova, Vopr. Radioélektron. Ser. 3: Detali Komponenty Apparat., No. 4, 3 (1963).
[7] S. Chikazumi, *Physics of Ferromagnetism, Vol. 2: Magnetic Characteristics and Engineering Application* [in Japanese], Syokabo, Tokyo (1978); *Physics of Ferromagnetism*, 2nd. ed., (Oxford University Press, New York, 1997, Mir, Moscow, 1987, 419 pp.).
[8] K. P. Belov and M. A. Zaĭtseva, in *Ferrites* [Russian translation from English], Moscow (1962), 504 pp.

Translated by P. Shelnitz

Kinetics of the failure of loaded materials at variable temperature

V. I. Betekhtin, V. M. Roĭtman, A. I. Slutsker, and A. G. Kadomtsev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

(Submitted September 5, 1997)
Zh. Tekh. Fiz. **68**, 76–81 (November 1998)

The possibility of predicting the lifetime of loaded materials at variable (increasing) temperature is demonstrated on the basis of a kinetic approach to the problem of the failure of such materials. © *1998 American Institute of Physics.* [S1063-7842(98)01511-6]

INTRODUCTION

Questions concerning the thermostability of materials and structures are of great importance for solving many problems in modern technology. If the temperature of materials or structures under a mechanical load should rise (in comparison to the normal service temperature), their strength begins to drop and failure can set in. Therefore, estimating the fitness of materials and predicting the possibility of failure in extreme situation associated with a rise in temperature (for example, in a fire) is an important and topical problem. The treatment of this problem has been the subject of numerous studies, predominantly taking a mechanical-engineering approach (see, for example, Refs. 1 and 2). For this reason, the solutions proposed have generally had a particular and narrowly specialized character, which, of course, does not detract from their practical significance. The treatment of this problem on the basis of modern physical theories regarding the failure of solids has just been started.[3,4]

A physical basis for analyzing the mechanical failure of bodies as the temperature varies is provided in a natural way by conceptions of failure as a kinetic phenomenon, rather than a critical act.[5,6] The macroscopic fracture of a loaded body is preceded by the formation and development of macroscopic cracks and pores, which develop in the body from the moment when the load is applied, and the actual fracture is the final act in this process.

The lifetime τ, i.e., the time from the moment when a load is applied until fracture of the body occurs, is an integral characteristic of this process. In the simplest case, in which the tensile stress σ and the temperature T are held constant over the course of the lifetime, the general expression obtained[5] for the lifetime of an extensive list of materials over a fairly broad range (of not excessively small or excessively large values of σ and T) is

$$\tau=\tau_0 \exp\left(\frac{U_0-\gamma\sigma}{kT}\right), \tag{1}$$

where $\tau_0 \approx 10^{-13}$ s is close to the vibration period of atoms, U_0 coincides with the interatomic bond dissociation energy, γ includes the activation volume and the local excess stress coefficient (a consequence of the structural inhomogeneity of the material), and k is Boltzmann's constant.

The basic form of expression (1) as a Boltzmann factor, the closeness of τ_0 to the vibrational period of atoms, and the identification of the initial barrier U_0 that is being removed by the applied stress σ permit one to characterize the kinetics of failure as a process controlled by the thermal-fluctuational decay of stressed interatomic bonds, which leads to the formation and development of microscopic cracks (defects) in the loaded body.

The analysis of the kinetics of failure under conditions where the temperature is held constant ($T=\text{const}$) but the stresses vary with time t (i.e., $\sigma=\sigma(t)$) has been the subject of several studies.[5,7] It was shown in these studies that the mechanism of failure does not vary (the values of τ_0, U_0, and γ remain unchanged) and that only the rate of failure varies with time. Formulas for calculating the lifetime under the conditions $\sigma=\sigma(t)$ at $T=\text{const}$ were derived.

No similar analysis has hitherto been performed for the case of $\sigma=\text{const}$ and $T=T(t)$. Just such a situation is characteristic of the case of a rising temperature, which is often realized in practice.

The purpose of the present work is to examine the kinetics of failure with variation of the temperature of a loaded body with time [i.e., under the conditions $\sigma=\text{const}$ and $T=T(t)$].

OBJECTS OF INVESTIGATION AND CONDITIONS

A typical construction material, viz., class A-1 reinforcement steel (GOST 5781-75), was chosen as the object of investigation. Samples were prepared in the form of rods with a diameter of 10 mm and a length of the working part equal to 400 mm. In addition, flat samples in the form of a double blade with a thickness of 0.2 mm and a length and width of the working part equal to 22 and 3 mm, respectively, were tested. The samples were tested under conditions of uniaxial extension. The lifetime was measured in two regimes: 1) under static conditions, i.e., constancy of $\sigma=\text{const}$ and $T=\text{const}$ during the fracture of each individual sample, with variation of σ and T from sample to sample; 2) with variation of the temperature for a sample loaded to the same value of σ. All the measurements were performed in an air environment.

STATIC LIFETIME

The results of measurements of $\tau\ (\sigma,T)$ curves under static conditions are presented in Fig. 1a. The plots in semilogarithmic coordinates comprise a family of straight lines, whose slopes decrease with increasing temperature. After ex-

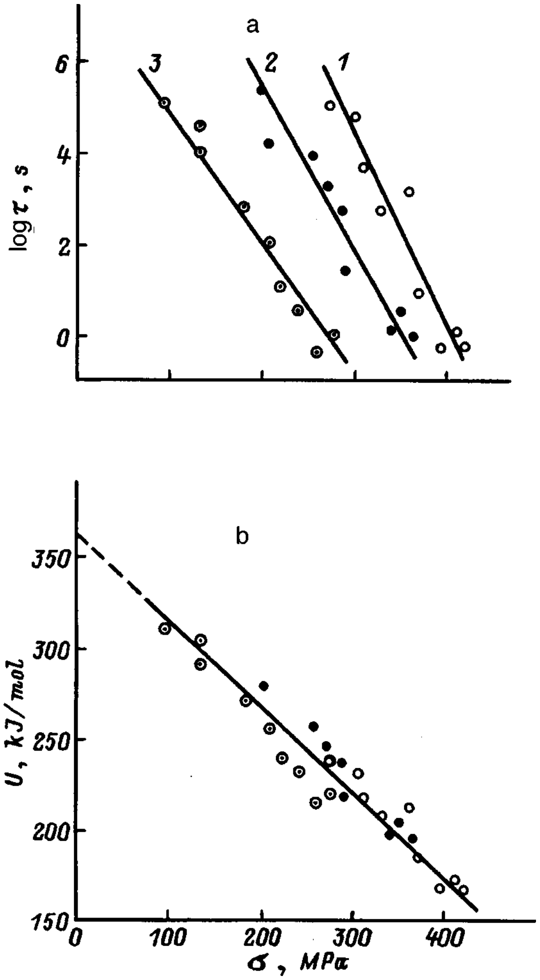

FIG. 1. a — Dependence of the lifetime on stress for steel samples at various temperatures: *1* — 673, *2* — 773, *3* — 873 K. b — dependence of the activation energy for the failure of steel on stress.

trapolation in the direction of increased values of σ, the straight lines converge to a hub lying near 10^{-13} s (i.e., the family of lines has a fan-like pattern).

Does the family of $\tau(\sigma,T)$ curves just described correspond to the general kinetic expression (1)? It follows from (1) that

$$U(\sigma)=U_0-\gamma\sigma=kT(\ln\tau-\ln\tau_0). \tag{2}$$

Each point in Fig. 1a was recalculated using Eq. (2). In accordance with the position of the hub in Fig. 1a, the value of τ_0 was set equal to 10^{-13} s. The results of the recalculation are shown in Fig. 1b, where the symbols corresponding to different temperatures are the same as in Fig. 1a. Fairly close bunching of all the points along a single straight line can be seen. This shows that the relation (1) holds for this steel. Thus, the kinetics of the failure of this construction material conform to the general physical laws governing the failure of solids.

The value of U_0 can be found from Fig. 1b (by extrapolation to $\sigma=0$): $U_0=360$ kJ/mol$=3.9$ eV. This value is fairly close to the dissociation (sublimation) energy of iron.[5]

The slope of the plot in Fig. 1b gives $\gamma=0.8$ nm^3. In accordance with the interpretation of the coefficient γ (see Refs. 5 and 6), $\gamma=q\cdot V_a$, where V_a is the activation volume in an elementary act of decay of interatomic bonds and has a typical value of 10^{-2} nm^3, and q is the local excess stress coefficient. We thus obtain $q=80$. Such a value for the local excess stress coefficient lies in the typical range of excess stresses for polycrystalline metals.[5]

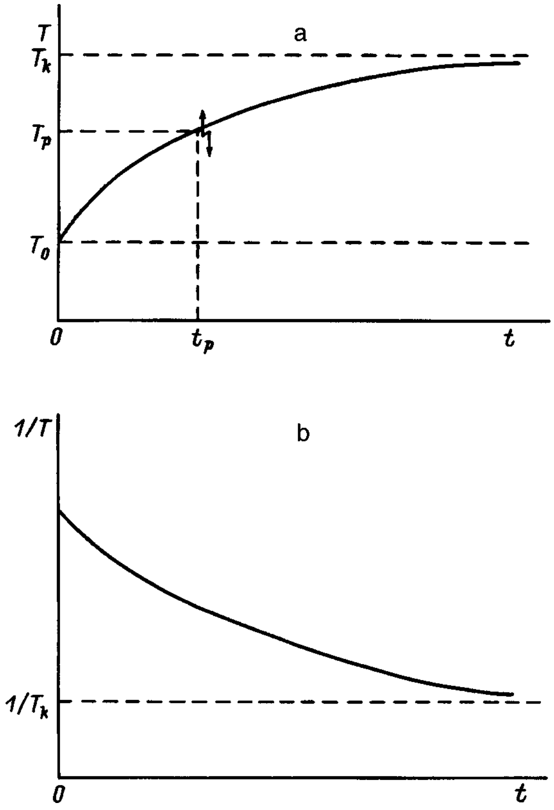

FIG. 2. a — Schematic representation of a monotonic increase in temperature with time corresponding to a normal fire regime; b — dependence of the reciprocal temperature on heating time.

Thus, the construction materials investigated conform completely to the fundamental assumptions of the kinetics of the failure of solids. This situation allows us to move on to a discussion of the laws governing failure in a regime with variation of the temperature during the time that the sample remains under load.

LIFETIME IN A REGIME WITH VARIATION OF THE TEMPERATURE

The variation of the temperature with time was assigned by the corresponding temperature increase during a ''normal'' fire.[4] A schematic representation of a typical monotonic rise in temperature is shown in Fig. 2a. The rate of increase in temperature slows with time, and T approaches a more or less pronounced asymptotic value T_k. For a sample under the influence of a certain constant tensile stress σ, fracture occurs at the time t_f at the value of the temperature at that moment T_f. Of course, the greater is σ, the smaller are the values of t_f and T_f.

Two ''normal'' fire regimes were assigned in the present work: regime 1 and regime 2, which differ with respect to the heating rate (or intensity). Plots of the rise in temperature

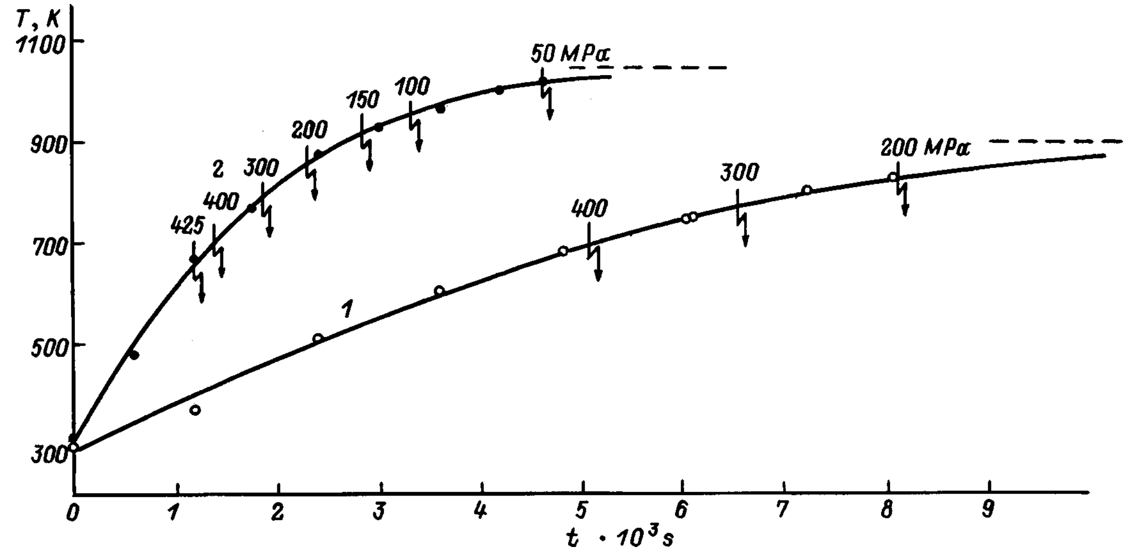

FIG. 3. Monotonic temperature-variation regimes for different heating rates. The arrows indicate fracturing of the samples under different stresses.

with time for the two regimes are shown in Fig. 3. It is seen that the common functional course of $T(t)$ corresponds to the form of the $T(t)$ curve schematically represented in Fig. 2a.

The arrows on the curves in Fig. 3 indicate fracturing of the samples under different stresses. Thus, a set of values of t_f and T_f corresponding to the values of the applied stress σ is obtained for each of the regimes. The values of t_f and T_f as functions of σ are listed in Table I.

The approach used in the present work to describe the lifetime of a loaded body under variation of the temperature with time is based on the simple and natural argument that defects accumulate in a loaded body during preparation for failure. This argument is based on numerous data on the processes occurring in loaded bodies.[5,6]

If the temperature of a loaded body varies with time, i.e., if $T=T(t)$, the principle of defect summation leads to Bailey's condition[8], i.e., a criterion for the failure of a body,

$$\int_0^{t_f} \frac{dt}{\tau[\sigma,T(t)]}=1, \tag{3}$$

where t_f is the time to failure.

TABLE I. Calculated and experimental values of the lifetime of loaded steel samples at increasing temperatures.

Regime	σ, MPa	T_f, K	$t_f\times10^3$ s	Calculation from data in Fig. 3: $t_f\times10^3$ s
1	200	820	8.1	10.7
	300	770	6.5	–
	400	690	5.1	5.9
2	50	1020	4.6	–
	100	950	3.3	4.4
	150	910	2.8	–
	200	860	2.3	–
	300	790	1.85	–
	400	690	1.40	1.3
	425	660	1.20	–

Expression (3) means that the ''partial'' lifetimes, i.e, the relative fractions of the lifetime at different values of T, are summed, and thus it corresponds to a summation of the defects accumulated at different temperatures.

It follows from (1) that $\tau[\sigma,T(t)]=\tau_0 \exp U(\sigma)/kT(t)$, and expression (3) becomes

$$\frac{1}{\tau_0}\int_0^{t_f} \exp\left[-\frac{U(\sigma)}{kT(t)}\right] dt=1. \tag{4}$$

Thus, the problem of calculating the integral in (4) for a specified form of the function $T(t)$ has been defined. The explicit form of the function $T(t)$, or the equivalent form of the function $1/T(t)$, is needed to obtain an analytic solution, and integrability of expression (4) is desirable for this purpose. Switching to $1/T(t)$ would, of course, be useful because the function is in the numerator, rather than the denominator, of the exponential function in the integrand, and this change will certainly facilitate finding the solution.

The form of the function $1/T(t)$ derived from the function $T(t)$ is schematically represented in Fig. 2b. Incidently, it is easier to mark the asymptotic level of $1/T_k$ using this plot. The plot of $1/T(t)$ resembles a decaying exponential function, and this calls for replotting the data in Fig. 3 on $T(t)$ as plots of the dependence of $\log(1/T-1/T_k)$ on t. The estimates of the level of T_k are indicated by the dashed lines in Fig. 3.

Figure 4 presents the results of such replotting. It can be seen that the plots of the dependence of $\log(1/T-1/T_k)$ on t for the two regimes are nearly linear.

Therefore, as a good approximation of the function $1/T(t)$ we can take

$$\frac{1}{T}(t)=\frac{1}{T_k}+\frac{1}{T_a}\exp(-\alpha t), \tag{5}$$

where $1/T_a=1/T_0-1/T_k$ (T_0 is the value of T at $t=0$; here $T_0=293$ K).

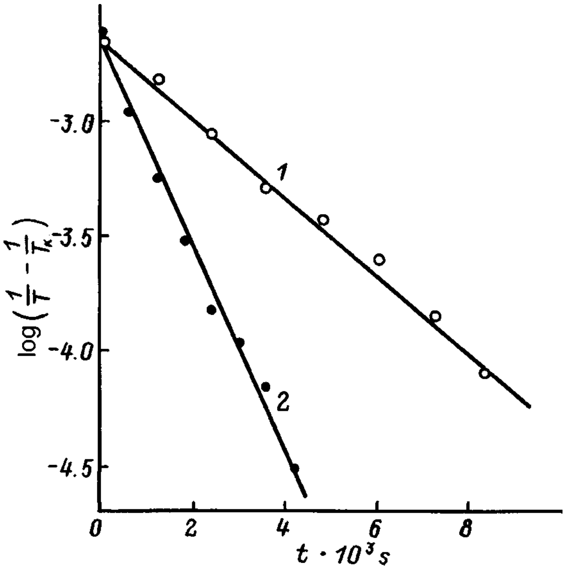

FIG. 4. Replotting of the data in Fig. 3 in new coordinates.

Then

$$\int_0^{t_f}\exp\left[-\frac{U(\sigma)}{kT(t)}\right]dt=\exp\left[-\frac{U(\sigma)}{kT_k}\right]\times\int_0^{t_k}\exp\left[-\frac{U(\sigma)}{kT_a}\exp(-\alpha t)\right]dt.$$

We make the change of variables $U(\sigma)/kT_a\ \exp(-\alpha t)=z$.

Then

$$\int_0^{t_f}\exp\left[-\frac{U(\sigma)}{kT_a}\exp(\alpha t)\right]dt$$

$$=-\frac{1}{\alpha}\int_{U(\sigma)/kT_a}^{U(\sigma)/kT_a\exp(-\alpha T_f)}\frac{\exp(-z)}{z}dz$$

$$=\frac{1}{\alpha}\left[\int_{U(\sigma)/kT_a\exp(-\alpha T_f)}^{\infty}\frac{\exp(-z)}{z}dz-\int_{U(\sigma)/kT_a}^{\infty}\frac{\exp(-z)}{z}dz\right].\tag{6}$$

The integrals in (6) cannot be taken exactly, but there is known approximation

$$\int_b^{\infty}\frac{\exp(-z)}{z}dz\simeq\frac{\exp(-b)}{b}\quad\text{for } b>1.\tag{7}$$

In order to be able to use this approximation, we must estimate the lower limits of the integrals in (6).

It follows from the data in Fig. 1b that the values of $U(\sigma)$ for stresses varying from 400 to 100 MPa lie in the range 200−300 kJ/mol. As follows from Fig. 4, $1/T_a$ equals $2.3\times10^{-3}\ \mathrm{K}^{-1}$. Then the lower limit of the second integral in (6) lies in the range 50–90, i.e., is much greater than unity.

According to the slopes of the plots in Fig. 4 and the expression (5), the values of α are $\alpha_1=0.39\times10^{-3}\ \mathrm{s}^{-1}$ for regime 1 and $\alpha_2=1.07\times10^{-3}\ \mathrm{s}^{-1}$ for regime 2. Then, taking the data on the failure time t_f from Fig. 3 for the corresponding stresses σ, we obtain a range of values from 3 to 7 for the lower limit of the first integral in (6) for regime 1 and a range from 3 to 10 for regime 2. Thus, in all cases the lower integration limits in (6) have values greater than 1, which permit the use of the approximation (7) for the integrals in (6).

Since, as follows from the estimates just given, the lower limits of the second integral in (6) are significantly higher than those of the first integral, then, according to (7), the second term in (6) is much (by an order of magnitude) smaller than the first, and the second term can thus be neglected. As a result, from (4) and (6) we obtain

$$\frac{1}{\alpha}\exp\left[-\frac{U(\sigma)}{kT_a}\exp(-\alpha t_f)\right]\left[\frac{U(\sigma)}{kT_a}\exp(-\alpha t_f)\right]^{-1}\simeq\tau_0\exp\left[\frac{U(\sigma)}{kT_k}\right].\tag{8}$$

We take into account that, in accordance with (5),

$$\frac{1}{T_f}=\frac{1}{T_k}+\frac{1}{T_a}\exp(-\alpha t_f).$$

Then, from (8) we obtain

$$U(\sigma)=kT_f\left[\alpha t_f-\ln\left(\tau_0\alpha\frac{U(\sigma)}{kT_a}\right)\right].\tag{9}$$

In (9) the dependence of the second term on the variables appearing in it α, $U(\sigma)$, and T_a is logarithmic, i.e., weak. With consideration of the relatively weak variability of these quantities $[\alpha=(0.4-1.0)\times10^{-3}\ \mathrm{s}^{-1}$, $U(\sigma)=200-300$ kJ/mol, and $1/T_a=2.3\times10^{-3}\ \mathrm{K}^{-1}$ is essentially constant] and the large value $\log\tau_0=-13$ of the control parameter, we can assume that $\ln(\tau_0\cdot\alpha\cdot U(\sigma)/kT_a)$ is approximately constant:

$$\ln\left(\tau_0\cdot\alpha\cdot\frac{U(\sigma)}{kT_a}\right)\approx-33.\tag{10}$$

Then

$$U(\sigma)\simeq kT_f(\alpha\cdot t_f+33).$$

Thus, an approximate expression has been obtained for determining the activation energy of the failure process from the values of the failure time t_f and the failure temperature T_f in an assigned temperature-variation regime, which is characterized by the coefficient α.

Expression (10) enables us to find the $U(\sigma)$ curve for the steel from the data plotted in Fig. 3 and to compare it with the $U(\sigma)$ curve obtained in the experiments with constant σ and T (Fig. 1b). The results of the determination of $U(\sigma)$ from the data in Fig. 3 are presented in Fig. 5. Fairly dense bunching of the points (corresponding to both temperature-variation regimes) along a straight line can be seen. The plot in Fig. 5 is close to the plot of $U(\sigma)$ in Fig. 1b. It is described by a function of the same form, i.e., $U(\sigma)=U_0-\gamma\sigma$. In addition, it is found from Fig. 5 that $U_0=340$ kJ/mol $=3.7$ eV and $\gamma=0.7\ \mathrm{nm}^3$, i.e., the values of U_0 and γ for the data in Figs. 1b and 5 are fairly close.

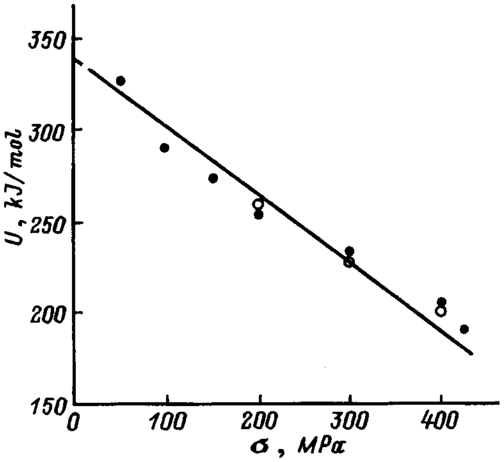

FIG. 5. Dependence of the activation energy for the failure of steel on stress according to the data in Fig. 3.

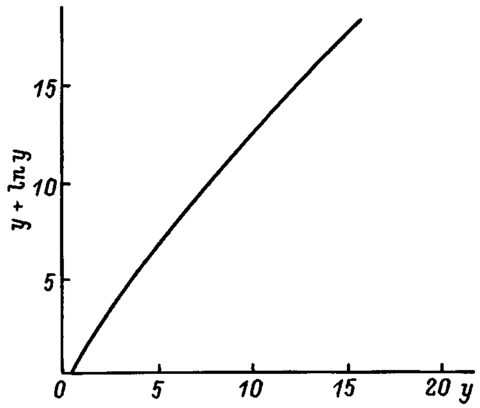

FIG. 6. Numerical plot of the dependence of $(y+\ln y)$ on y.

This means that the nature of the failure process at a temperature which varies with time remains both qualitatively and quantitatively the same as that which has been reliably revealed in experiments with constant σ and T. With this finding one can develop methods for predicting the lifetime of materials or structures at varying temperature on the basis of the general physical laws governing the kinetics of failure.

Implementation of this possibility will require: 1) knowledge of the dependence of the temperature on time, i.e., $T(t)$, and 2) knowledge of the kinetic characteristics of the failure of the material under consideration, i.e., the values of U_0, γ, and τ_0. These characteristics can be obtained by measuring the dependences of the lifetime on stress σ and temperature T, as was shown above (Fig. 1).

We now present examples of estimates of the time to failure (t_f) for the steel investigated in the present work in the same heating regimes, using the values of $U(\sigma)$ from Fig. 1 for specified values of σ.

Such an estimate was made on the basis of relation (8). The complicated form of this expression precludes a direct analytical solution with a determination of t_f from the specified values of σ. Therefore, we make the substitution

$$\frac{U(\sigma)}{kT_a}\exp(-\alpha t_f)=y. \tag{11}$$

Then, from (8) we have

$$\exp(-y)/y=\alpha\tau_0\exp\left[\frac{U(\sigma)}{kT_k}\right]. \tag{12}$$

Taking the logarithm of (11), we obtain

$$y+\ln y=-\ln\left\{\alpha\tau_0\exp\left[\frac{U(\sigma)}{kT_k}\right]\right\}. \tag{13}$$

The right-hand side of (13) contains quantities known from Figs. 1 and 2 [we assign σ and determine $U(\sigma)$ from Fig. 1b]. We numerically construct the dependence of $y+\ln y$ on y (Fig. 6) and then use it to find the values of y from the values of the right-hand side of (13). Now we can use them to find the values of t_f, i.e., the values of the time to failure for different values of the stress σ from (11). Estimates were made for regime 1 with $\sigma=200$ and 400 MPa and for regime 2 with $\sigma=100$ and 400 MPa. The calculated values of t_f are listed in Table I, where they can be compared with the experimentally obtained values of t_f for the same stresses.

It can be seen that the calculated values differ only slightly from the measured values. Thus, the satisfactory reliability of predictions of the time to failure has been demonstrated for reinforcement steel. We note that similar estimates, which confirmed the legitimacy of a similar approach to predicting the lifetime at increasing temperature, were obtained in Ref. 9 for samples of concrete, whose temperature-force dependences closely conform to expression (1).

The problem of estimating the safe stress for an assigned temperature-increase time can also be solved on the basis of (11). Of course, cases of different forms of the function $T(t)$ can also be of interest. Then other methods, both analytical and numerical, for calculating integrals of the type (4) are needed (see, for example, Ref. 5). However, the physical principles of the approach to solving these problems should clearly still be based on the kinetic concepts of the failure of solids.

[1] N. A. Strel'chuk, *Explosionproofing and Fire Resistance in Construction* [in Russian], Stroĭizdat, Moscow (1970), 271 pp.

[2] O. Pettersson, Fire Saf. J. **13**(1), 1 (1988).

[3] V. M. Roĭtman, V. I. Betekhtin, and N. A. Parfent'eva *et al.*, Izv. Vyssh. Uchebn. Zaved. Stroit. Arkhit., No. 11, 61 (1983).

[4] V. M. Roitman, J. Appl. Fire Sci. **3**(1), 3 (1993).

[5] V. R. Regel', A. I. Slutsker, and É. E. Tomashevskiĭ, *Kinetic Nature of the Strength of Solids* [in Russian], Nauka, Moscow (1974), 573 pp.

[6] P. G. Cheremskoĭ, V. V. Slezov, and V. I. Betekhtin, *Pores in Solids* [in Russian], Énergoatomizdat, Moscow (1990), 376 pp.

[7] S. N. Zhurkov and É. E. Tomashevskiĭ, in *Some Problems in the Strength of Solids* [in Russian], Leningrad (1959), pp. 68–73.

[8] J. Bailey, Glass Ind. **20**, 21 (1939).

[9] V. I. Betekhtin, V. S. Kuksenko, A. I. Slutsker, and I. É. Shkol'nik, Fiz. Tverd. Tela. (St. Petersburg) **36**, 2599 (1994) [Phys. Solid State **36**, 1416 (1994)].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

On the dimensionality of superconductivity in cuprate superconductors

M. V. Krasin'kova

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

(Submitted April 15, 1998)
Zh. Tekh. Fiz. **68**, 82–84 (November 1998)

The question of the dimensionality of superconductivity is considered within the framework of a model of superconductivity via asymmetric, delocalized ''crystalline'' π orbitals (analogous to the corresponding molecular orbitals) extending along chains of covalently bonded copper and oxygen ions. It is shown that superconductivity is preceded by a separation of the bonds in the CuO_2 layer into covalent and ionic bonds with ordering of the covalent bonds into chains. Such an ordering facilitates the formation of a crystalline π orbital lowering the crystal energy by the resonance energy of the π bond and is therefore favored. The superconducting current is created by non-dissipative motion of π-electron pairs along the asymmetric, ''crystalline'' π orbitals extending along chains of covalently bonded copper and oxygen ions, in the presence of an ionic bond between neighboring chains extending through the easily polarizable O^{2-} ions. This ionic bond correlates the motion of the electron pairs along all the π orbitals and stabilizes the superconducting state. Only in this sense is the apparent ''one-dimensionality'' of superconductivity in cuprate superconductors to be understood. *© 1998 American Institute of Physics.* [S1063-7842(98)01611-0]

A model was proposed in Ref. 1 which explained the nature of high-temperature superconductivity in cuprate superconductors by proceeding from a picture of electron pairs in the material forming a chemical bond (an asymmetric π bond) which under certain conditions becomes delocalized and enables superconductivity.

According to this model, the superconducting pairs move along spatially separated, asymmetric π orbitals extending along chains of covalently bonded copper and oxygen ions lying in the CuO_2 plane. The latter circumstance creates the impression that the superconductivity is one-dimensional.

Let us examine the problem of the dimensionality of the superconductivity of cuprate superconductors in this model in more detail and consider what may be the cause of the formation of these chains of covalently bonded ions, why a superconducting current is possible only along chains lying in one plane, and how such apparently one-dimensional superconductivity is stabilized.

The difference in the electronegativities of copper and oxygen is such that the bond in the CuO_2 layer should be half ionic and half covalent. Let us see how this condition is fulfilled in these cuprate structures. The coordination polyhedron of copper in most cuprate superconductors is a square-based pyramid. Thus, the copper ion has five nearest neighbors and should therefore form at least five bonds.$^{1)}$

Jumping somewhat ahead, note that copper forms not five, but six bonds: three ionic bonds (two in the CuO_2 plane and one along the C axis) and three covalent bonds $(2\sigma+1\pi)$. But such a state of the bonds, in complete agreement with the difference in electronegativities and therefore stable, is apparently reached only in the superconducting state.[1]

In the undoped material, the achievement of such a balance between ionicity and covalency of the bonds as a result of the difference in electronegativities runs up against a series of difficulties. First of all, the presence of a flat-square coordination of the Cu^{2+} ion requires that four bonds in the CuO_2 plane be equivalent, which is quite a few for just one type of bond. Second, in the presence of an unpaired electron on the Cu^{2+} ion and a pair of electrons on the O^{2-} ion in orbitals perpendicular to the CuO_2 plane, the formation of a π bond between these ions becomes impossible. Third, the distance between the Cu^{2+} and O^{2-} ions is greater than the sum of their covalent radii and increases with growth of the temperature as a result of thermal expansion of the lattice, which also hinders the formation of a covalent bond.

A way out of this predicament for the undoped material is provided by the state of covalent–ionic resonance.[2] The term ''state of covalent–ionic resonance'' means that the bond between the ions is found in some intermediate state — neither ionic nor covalent.[3] Therefore, the distance between the ions can also have some intermediate value, i.e., it can be less than the value expected for the ionic bond, but greater than the value required for the covalent bond, as is in fact observed experimentally.

It is hard to say how closely the required balance between ionicity and covalency of the bond in the copper coordination polyhedron is fulfilled under conditions of covalent–ionic resonance of four bonds in the CuO_2 plane and one ionic bond along the C axis. It is more likely that it is not fulfilled completely and the bond is not completely stable, although this instability is partly compensated by a lowering of the energy of the system by the energy of the covalent–ionic resonance. But when the material is doped or changes are made in its stoichiometry or additional layers are introduced incommensurate with the copper layers, the insta-

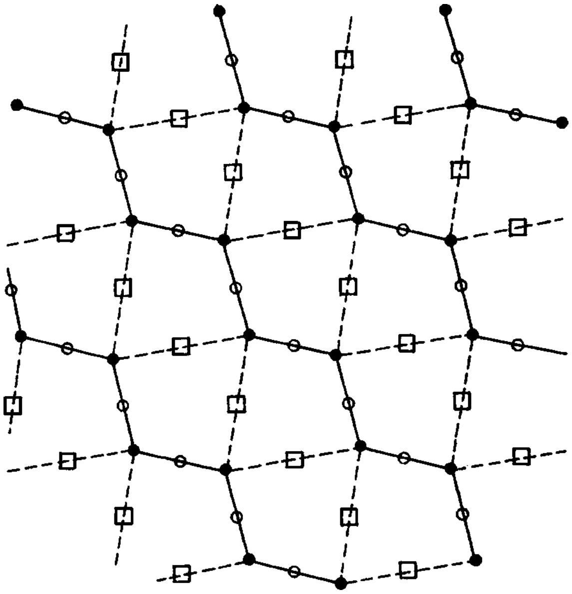

FIG. 1. Diagram of the CuO_2 plane. The chains of covalent bonds are represented by the heavy lines joining copper ions (filled circles) and oxygen ions (unfilled circles). The dashed lines are ionic bonds between chains extending through the O^{2-} ions (squares). Superconducting pairs move along asymmetric π orbitals extending along the chains and located on one side of the CuO_2 plane (above or below the plane). The p_z orbital of the O^{2-} ion bridge, deformed by the field and containing two electrons, is also found on that side of the plane.

bility of the bond manifests itself immediately in the ease with which the flat-square coordination of the copper ions is distorted. If upon doping it becomes possible for an exchange of electrons between the CuO_2 layer and the doped layers to occur, then the bonds in the CuO_2 plane will separate into two ionic bonds (with greater distance between the ions) and two covalent σ bonds (with a smaller distance between the ions). To achieve the required balance between ionicity and covalency of the bonds in this case, in addition to the two covalent bonds just mentioned, a π bond is also formed. This π bond is formed by the overlap of copper and oxygen orbitals perpendicular to the CuO_2 plane that between themselves have lost one electron to doping (the electron goes into the doped layer).2)

But separation of the bonds in the CuO_2 plane into two ionic and two covalent bonds (with one of the latter a double bond) takes place apparently not in a random fashion, but in an ordered fashion — with the formation of chains of covalently bonded ions. These chains turn out to be ionically bonded with each other through the O^{2-} ions (see Fig. 1). Such ordering is energetically favorable for the system as a consequence of the possibility of lowering the energy of the system by the resonance energy of the π bond along the entire length of the chains.3)

On the other hand, ordering of the covalent bonds in the form of chains by itself creates conditions for the formation of a ''crystalline'' π orbital (analogous to the corresponding molecular orbital) extending along the entire length of the chains of covalently bonded ions and necessary for the emergence of the superconducting state, as will become apparent below.

Recall that we are talking here about an asymmetric π orbital. The asymmetry is created by the asymmetric electric field on either side of the CuO_2 plane. The orbitals of the copper and oxygen ions perpendicular to the CuO_2 plane are deformed by this asymmetric field — the ''volume'' of their lobes increases, facilitating a better overlap on the side of the CuO_2 plane facing the larger positive charge.[1]

Thus, from the above consideration it follows that the formation of ''crystalline'' π orbitals running along chains of covalently bonded ions is a consequence of the mixed covalent–ionic character of the interaction between copper and oxygen and is energetically favorable for the system since it lowers its energy by the resonance energy of the π bond.

Delocalization of the π electrons of the asymmetric π bond along the chains in fact implies the possibility of nondissipative motion of electron pairs along a ''crystalline'' π orbital. The chains of copper and oxygen ions with delocalized π electrons in this case reminds one of conjugated bonds in molecules of aromatic compounds with the only difference being that they are unclosed, and their π orbital is asymmetric about the plane of the ''molecule'' (in the given case the CuO_2 plane). Such motion of the electron pairs along π orbitals constitutes motion along spatially isolated ''filaments'' located on one side of the CuO_2 plane, a circumstance which creates the impression of one-dimensionality.

But a necessary condition for the superconducting state, being a macroscopic phenomenon, is that the motion of the electron pairs on each separate chain and all of them together in the CuO_2 plane be correlated.

Correlation of the motion of the electron pairs along each separate chain is ensured by the strong repulsion between the electron pairs. As they move along a π orbital, they do not collide with each other but rather keep their distance as they move, maintaining a separation between them equal to the distance between the copper and oxygen ions in the chain.

Correlation of the motion of the electron pairs in neighboring chains is ensured by the ionic bond between them through the easily polarized O^{2-} ions, forming bridges between the chains. Here each copper ion of one chain is coupled through the two nearest O^{2-} ions with two copper ions of the neighboring chain (see Fig. 1). Under these conditions, polarization of the O^{2-} ion can transfer to a neighboring chain an excitation in the electron density distribution created by the motion of an electron pair along a π orbital in one of the chains. Thus, it is possible to realize correlated motion of electron pairs over all of the π orbitals extending along the chains in the CuO_2 plane.

It is necessary to allow for the fact that the O^{2-} ion also finds itself in an asymmetric field, and its orbital perpendicular to the CuO_2 plane and containing two electrons is also asymmetric and can overlap with the p_z orbitals of the copper ions joined by the given oxygen ion. However, this pair of electrons of the O^{2-} ion probably does not participate in the superconducting current along the chains, since it does not conform to the long-range order underlying the conjugation of the chain of covalently bonded ions, but it can apparently take part in screening currents.

In view of the O^{2-} bridge ions correlating the motion of the electron pairs, the superconductivity is not really one-dimensional even though the electron pairs move along spatially separated π orbitals.

As for the stability of superconductivity along the π orbitals, it is vouchsafed, as was seen above, by the lowering of the crystal energy by the resonance state of the π bond, correlated over all the chains in the CuO_2 plane.

Thus, the ordering of covalent bonds in the CuO_2 plane in the form of chains ionically bonded to each other ensures coherence of oscillations of the π-electron density over the entire CuO_2 plane.

Since the energy of the resonance state of the bond exceeds the energy of the electron–phonon interaction, it stands to reason that the the T_c of such superconductors will be higher than the T_c of superconductors with a BCS mechanism.

In conclusion, it should be said that some confirmation of the importance for superconductivity of the presence of chains of bonds may be garnered from the fact that in many presently known superconductors some aspects of one-dimensionality, or more precisely, a low dimensionality are observed. All complex materials in which superconductivity has been observed are layered, and in many of them metal–oxygen or metal–metal chains have been observed. On the other hand, it is interesting to note that in organic materials, conjugation of the bonds is observed only in molecules with a planar configuration.

Recently published observations of asymmetry in the oxygen environment of Cu in the CuO_2 plane (observation of two different Cu–O distances)[4] along with data pointing to localization of holes in the CuO_2 plane and the absence of any influence of these holes on the mobile carriers[5] not only do not contradict the above ideas, but can also be taken as a direct experimental confirmation.

[1)] Recall that to form covalent bonds on the Cu^{2+} ion, one $4s$ orbital, three $4p$ orbitals, and one half-filled $3d$ orbital are available. All these orbit-als are similar in energy. The O^{2-} ion has completely filled $2s$ and $2p$ orbitals.

[2)] In this case it is apparently not even important which of the ions loses the electron: the copper ion loses its unpaired electron and transitions from the Cu^{2+} state to the Cu^{3+} state, now having a free p_z orbital which now overlaps with a completely filled p_z orbital of an O^{2-} ion to form a π bond, or an oxygen ion transitions from the O^{2-} state to the O^{-} state. In the latter case, the p_z orbitals of the copper and oxygen ions can overlap, each one occupied by one electron. It may be supposed that upon doping, one electron is lost for every pair of electrons Cu^{2+}–O^{2-} forming between themselves a covalent bond. Note that in this case only one π bond is formed for every two σ bonds Cu–O–Cu.

[3)] The loss of the covalent–ionic resonance energy of four bonds upon the above-mentioned separation of the bonds appears to be compensated by the resonance energy of the π bond formed along the chains.

---

[1] M. V. Krasin'kova, Pis'ma Zh. Tekh. Fiz. **23**(17), 57 (1997) [Tech. Phys. Lett. **23**, 681 (1997)].

[2] L. Pauling, Phys. Rev. Lett. **59**, 225 (1987).

[3] L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals*, 2nd ed. (Cornell University Press, Ithaca, New York, 1940, Goskhimizdat, Moscow 1947, 440 pp).

[4] H. Oyanagi and J. J. Zegenhagen, Superconductivity **10**, 415 (1997).

[5] P. C. Hammel, B. W. Statt, R. L. Martin *et al.*, Phys. Rev. B **57**, R712 (1998).

Translated by Paul F. Schippnick

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Aspects of the self-organization of carbonaceous conducting nanostructures during electroforming of a metal–insulator–metal open sandwich structure with a nanometer-size insulating gap

V. M. Mordvintsev, S. E. Kudryavtsev, and V. L. Levin

Institute of Microelectronics, Russian Academy of Sciences, 150007 Yaroslavl', Russia

(Submitted June 24, 1997)
Zh. Tekh. Fiz. **68**, 85–93 (November 1998)

Experimental results are presented on the electroforming of a nanometer-size MIM (metal–insulator–metal) diode with a carbonaceous active medium. The diode is in the form of an MIM sandwich structure which is open on one face and has a nanometer-size insulating gap. Measurements of its current–voltage characteristics are made which reflect processes of self-organization and self-forming of carbonaceous conducting nanostructures in the insulating gap. It is shown that the properties of such a circuit element differ greatly from those of a conventional MIM diode. These differences can be explained if it is taken into account that a thin insulating layer is built in, in series with the carbonaceous conducting medium growing in the insulating gap. The data obtained indicate that the carbonaceous structure is of nanometer size in all three spatial dimensions. The models that have been developed to represent this structure correspond well with the experimental results, in particular the spatiotemporal self-organization in this system. © *1998 American Institute of Physics.* [S1063-7842(98)01711-5]

INTRODUCTION

Diodes with a metal–insulator–metal (MIM) structure, placed in vacuum with vapors of organic compounds and subjected to electroforming (as a rule, the cyclic application of a voltage ranging from 0 to 15 V), have been studied for quite some time now.[1] The insulating gap of a planar MIM structure, openly accessible to the organic vapor and having a characteristic width of the order of a micron, is prepared by the ''burnout'' of a thin metallic conductor located on the surface of an insulating substrate by means of an electric current. In the case of a sandwich MIM structure, it is created in a process similar to breakdown of the corresponding insulating film placed between metal electrodes. In both variants, the prepared structures possess a small-to-moderate initial conductivity, which increases by several orders of magnitude during electroforming. The *N*-shaped character of the quasistationary current–voltage (I–V) characteristics observed after electroforming have been linked with the formation of carbonaceous conducting pathways out of the adsorbed organic molecules, their burnout upon the passage of a certain level of current through the structures, and some process of regeneration when the voltage is decreased. We will refer to such structures as conventional MIM diodes. It has been experimentally shown that in their composition and structure the carbonaceous conducting pathways are similar to graphite.[2] Later it was found[3] that in such a formed MIM diode the entire applied voltage falls across the narrow insulating gap of nanometer width. The authors of Ref. 4, on the basis of scanning tunneling microscope experiments and the available data on conventional MIM diodes, proposed a mechanism of their functioning, the main elements of which are the nanometer width of the insulating gap and the presence of an internal reverse bias in such objects. This mechanism has been used previously to develop models for the appearance of *N*-shaped I–V characteristics in similar structures,[5,6] and the analysis of these models has led to an understanding of electroforming as a process of self-organization of a nanometer-size gap in the carbonaceous conducting medium.[7] To start up this process it is necessary to ensure that a flux of electrons can be passed through the organic material, which is initially an insulator. In a conventional MIM diode the necessary initial conductivity of the structure is provided by a special technology used in the formation of the insulating gap, which, after the conductor is burned out, is left with a modified surface and contains some residual metal particles. However, it was suggested back in Ref. 4 that the necessary conditions can be provided simply by making the nanometer-width initial insulating gap openly accessible to the organic molecules; then the high electric fields created in the organic medium on the surface of the gap will be sufficient for the appearance of a field-emission current from the cathode. Such an element will be significantly more controllable than a conventional MIM diode, while possessing substantially greater reproducibility of the spatial structure during fabrication. The passage of a flux of electrons through the organic material leads to dissociation of its molecules, carbon enrichment, and thus to the formation of a carbonaceous conducting medium (CCM) upon reaching the percolation threshold. In this paper we present experimental results confirming the validity of these conclusions. Such a structure has been called[5] a nano-MIM diode with a carbonaceous active medium.

SAMPLES AND EXPERIMENTAL TECHNIQUE

A nano-MIM diode with the width of the insulating gap not exceeding a few tens of nanometers, as is necessary in order to create high electric fields at the voltages actually used in electroforming, was created in the form of an open sandwich structure.[8] The insulating gap is an end face of the insulating film which is made accessible to the organic molecules by a local etching of the upper metal layer and the insulator layer of a typical three-layer MIM structure. Here, the gap width *H* is determined by the thickness of the insulating film, which can be prepared to an acceptable level of quality without any particular technical difficulties at thicknesses down to a few nanometers.

In the preparation of the samples, silicon wafers coated with silicon oxide served as the substrate. Metallic films with a thickness of $0.2-0.3\ \mu$m were deposited by magnetron sputtering. The material of the upper electrode (cathode) of the MIM structure was tungsten. The insulating film of aluminum oxide, which is one of the better insulators in the thin-film state, was formed by liquid anodization of the lower aluminum electrode. As is well known,[9] in this case a high-quality insulating layer is obtained which itself does not undergo electroforming in the sandwich structures. In our case this was a useful property, precluding electroforming beyond the limits of the open face of the insulating film. This face was created by precision etching of the aluminum oxide after local removal of the material of the upper electrode. The thicknesses *H* ranged from 15 to 40 nm, and the breakdown field strengths of the anodic aluminum oxide were around 9×10^6 V/cm. In a plane parallel to the substrate the structures had the form of crossed metallic bars separated by the anodic oxide. In the intersection region a square window was etched into the upper bar, through which the oxide was etched further. The dimensions of the window varied from 3 to 16 μm. The prepared structures were encased and placed in a vacuum chamber, which was pumped out in succession by a mechanical pump and then by a oil diffusion pump, thereby providing the usual gaseous medium for electroforming, containing vapors of organic molecules.

Figure 1 shows in block form one of the circuits that was used for electroforming and measurement of current–voltage characteristics. A triangular voltage waveform with amplitude up to 20 V and sweep rate of the order of a few V/s was imposed on the sample. Such a slow rate of change of the voltage ensured quasistationary I–V characteristics (in particular, ruling out effects arising due to an insufficiency of organic molecules adsorbed on the surface of the insulating gap), thereby making the results of the measurements independent of the parameters of the adsorption stage. A ballast resistance R_b was connected in series with the sample to lower the probability of catastrophic breakdown of the structure during electroforming (it later turned out that its role was significantly more important). The current was measured with a standard chart recorder.

Judging from the dimensions of the structures, primarily from the perimeter of the open face of the insulating film, and data on conventional MIM diodes, one may expect currents in the formed samples from a few to a few tens of milliamperes. However, the very first experiments showed that, even in the presence of a large ballistic resistance and with measures taken to eliminate parasitic capacitances in parallel with the sample, a catastrophic breakdown of the structure develops at significantly lower currents, shorting out the metal electrodes. Nevertheless, a more detailed observation in the range of currents 3–4 orders of magnitude less than expected showed that a process takes place in this case very similar to ordinary electroforming.

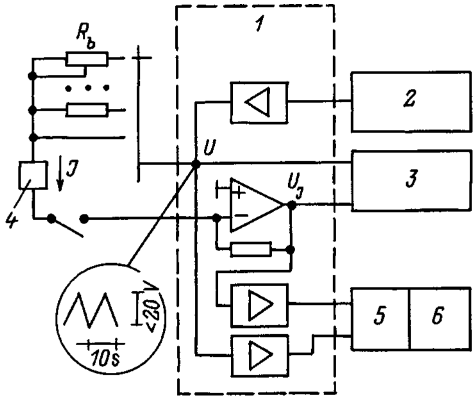

FIG. 1. Block diagram of the setup for electroforming and measurement of current–voltage characteristics: *1* — transformer unit, *2* — generator, *3* — chart recorder, *4* — sample, *5* — analog-to-digital converter, *6* — computer.

RESULTS AND DISCUSSION

Figure 2 plots typical curves characterizing the development of the current through an open MIM sandwich structure with an insulating gap about 20 nm in width during several cycles of variation (passes) of the generator voltage *U* at a constant ballast resistance. In the first pass the current is absent until $U=9$ V, where it appears with a jump. We call this point the breakaway point. With each successive pass, the current gradually grows and, for a given value of the ballast resistance, a limiting curve *4* is set up, which varies very little if *U* remains below a certain value. If this latter condition is violated, an abrupt (discontinuous) drop in the current is possible, with subsequent gradual recovery to its

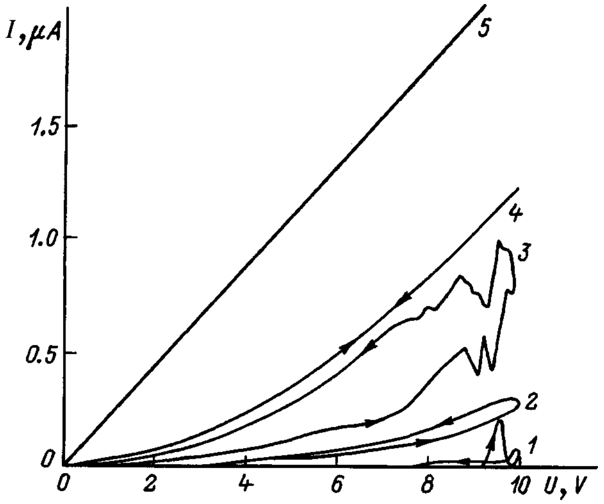

FIG. 2. Current–voltage characteristics of an open MIM sandwich structure, reflecting its electroforming: $R_b=5$ MΩ; *1* — first, *2* — third, *3* — fifth pass, *4* — after the tenth pass; *5* — I–V characteristic of the ballast resistance without the structure.

initial value. This limiting current–voltage characteristic depends on the magnitude of the ballast resistance, whose role consists in an automatic limitation of the growth of the voltage across the structure as the current grows in the circuit. For comparison, a straight line, corresponding to having only the ballast resistance in the circuit, is also plotted in Fig. 2. It can be seen that at large voltages the slope of the limiting curve *4* almost coincides with the slope of the straight line. This implies that in the given region, despite the growth in U, the voltage across the structure itself remains nearly constant. When $R_b = 0$ or its value is too small, a gradual growth of the current leads to catastrophic breakdown if the voltage amplitude U is not decreased accordingly.

The difference between such a process of electroforming and the usual process observed in conventional planar MIM diodes, and it is substantial, is that the initial nonzero conductivity appears only at some quite high voltage. But this is the way a nano-MIM diode should behave: the field emission from the cathode depends very abruptly on the field strength in the insulating gap. But after a current begins to flow through it, the formation of particles of the carbonaceous conducting phase begins, accompanied by formation of a carbonaceous conducting medium which gradually fills the insulating gap, thereby narrowing the residual clearance of the gap, which leads to growth of the conductivity of the structure.

Figure 3a shows the current–voltage characteristic of a sample that had been formed at a ballast resistance of $R_b = 5\ \text{M}\Omega$ to the limiting curve *4* of Fig. 2, after which the ballast resistance was bypassed. For comparison, Fig. 3b plots a typical I–V characteristic of a conventional formed MIM diode of planar design, with a width of the conducting bar of the same order as the perimeter of the face of the open sandwich structure (40 μm). It was fabricated by the burning out of a thin (15 nm) gold film and was formed in the same vacuum chamber. The similarity of the shape of the curves speaks of a similarity of the processes taking place in both cases, but there are important differences. First, the currents differ by 3.5 orders of magnitude for almost the same ''working'' perimeter of the structures. Second, an *N*-shaped curve is traced out in a conventional MIM diode on the reverse leg of the voltage cycle as well as on the forward leg, although with significant hysteresis. In the open sandwich structure this is not the case. Moreover, after being ''diverted,'' during the growth of U, to large values corresponding to complete shutoff, it remains, as a rule, on the *Off* branch for any number of subsequent passes. Third, in the case of a conventional MIM diode found on a branch of the current–voltage characteristic with negative differential resistance, after having been switched off by abruptly dropping the voltage across the structure to zero it also passes into to the zero-current state [the *Off* curve in Fig. 3a]. But this state is maintained only up to some threshold voltage U_{th} (around 2–2.5 V). With subsequent increase of the voltage, the device switches on (transitions to the *On* state); such behavior is entirely normal. For the open sandwich structure, on the other hand, in order to get it to switch on (transition to the *On* state) it is necessary to apply a voltage to it several volts greater than the switch-off amplitude, having first connected the ballast resistance in the circuit. In this case, a process develops that is similar to accelerated forming, as opposed to the slower, unaccelerated forming (Fig. 2). It appears as if the threshold voltages U_{th} are significantly greater and exceed the sweep amplitude for a switching-off pass.

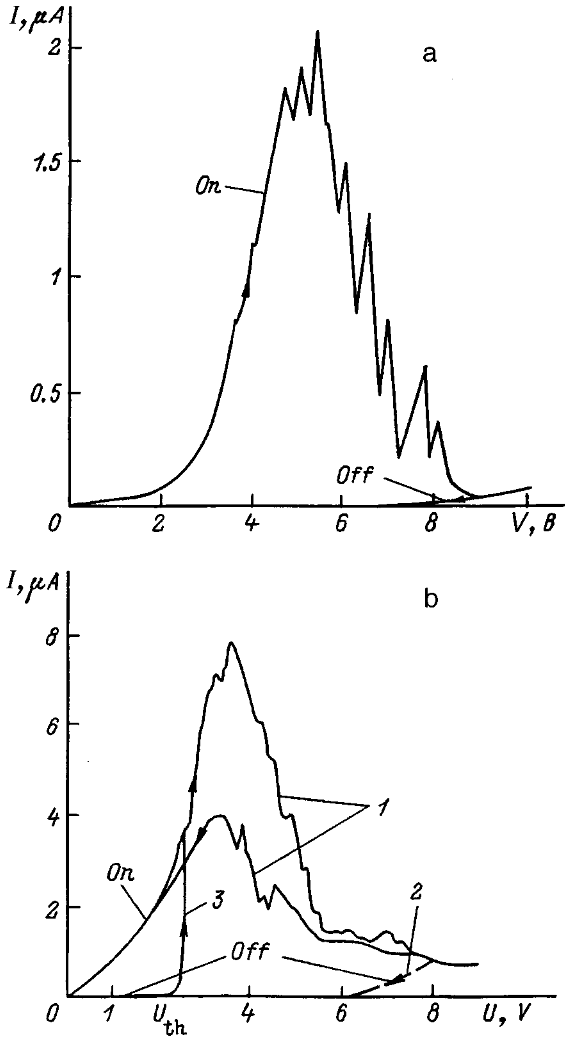

FIG. 3. Current–voltage characteristics of an open MIM sandwich structure for $H = 24$ nm, $R_b = 0$ (a) and for a conventional planar MIM diode for a 20 μm bar (b): *1* — sweep analogous to that shown in part a, *2* — voltage rapidly switched off at $U = 8$ V, *3* — sweep analogous to case *1* after rapidly switching off the voltage at $U = 8$ V.

The lower currents in an open sandwich structure can simply be attributed to the circumstance that a relatively small part of its ''working'' perimeter undergoes forming. However, if the carbonaceous formed structure on even a small segment of the perimeter were entirely equivalent to a conventional MIM diode, one would also observe an *N*-shaped I–V characteristic on the reverse leg of the voltage cycle. The absence of current growth with decreasing voltage implies that some factor exists that limits regeneration of the carbonaceous conducting medium. But since the reason for its formation, as is clear from the model devised in Ref. 5, is an electron flux through the adsorbed organic material, the influence of this factor may consist simply in the fact that it limits the current density in the structure, i.e., not only are the currents small, but the local current densities in the insu-

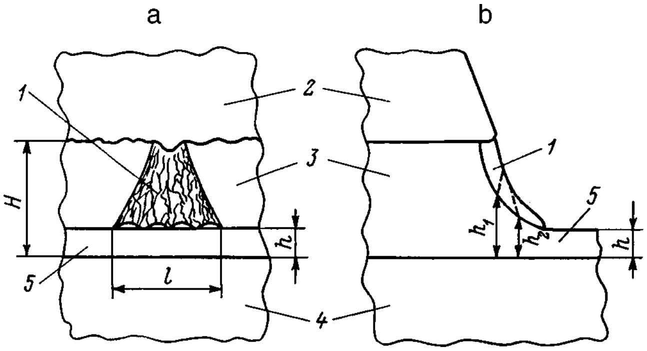

FIG. 4. Diagram of a nanometer-size insulating gap of an open MIM sandwich structure showing the carbonaceous conducting structure *1* that is formed in the electroforming process: a — end view of the open face, b — profile of the structure; *2* — metal cathode, *3* — insulating aluminum-oxide film, *4* — metal anode, *5* — residual oxide.

lating gap are also less than their usual values. On the other hand, if the current densities in the structure are significantly lower, it is not clear why burnout of segments of the CCM begins in practically the same voltage range as in a conventional MIM diode, this process being reflected by the position of the maximum and of the descending branch of the I–V characteristic in Fig. 3a.

All of the noted peculiarities of the open MIM sandwich structure can be explained if we take into account the presence of a residual oxide on the surface of the aluminum electrode after the anodic oxide is etched away (Fig. 4). Even if the latter were completely removed, a contiguous layer of the native oxide would quickly form on the aluminum surface, whose thickness (not less than 1.7 nm) would exceed the width of the auto-formed insulating gap in the CCM of a conventional MIM diode (about 1 nm; Refs. 3 and 6). Thus, in the insulating gap a thin layer of insulator material is connected in series with the CCM and limits the current density in the structure.

In this case the mechanism of electroforming and conduction in the device (in view of the results of the model proposed and developed in Refs. 6 and 7) looks as follows. As the voltage on the structure is increased, field emission begins when the breakaway point is reached (Fig. 2) from some nanoprotrusion on the surface of the cathode (Fig. 4a). Electrons injected into the organic insulator (the adsorbate layer) cause the molecules of the organic material to dissociate, resulting in the formation of particles of the carbonaceous conducting phase. Due to divergence of the electron flux, the current density falls with distance from the cathode, but since it is small in the initial stages of electroforming, the temperatures caused by Joule heating do not reach the values necessary for disappearance of the particles of the carbonaceous conducting phase[4,6] even near the cathode. Therefore, formation of a connected conducting cluster of carbonaceous particles (percolation) begins, in contrast to the case of a conventional MIM diode, on the cathode side, i.e., where the current density is maximum. Thus the CCM is initiated near the metal cathode, and as it forms it propagates toward the anode, decreasing the width of the insulating gap and increasing the conductivity of the structure, as is manifested in a gradual increase in the current (Fig. 2). The carbonaceous conducting medium in this case plays the role of an effective cathode. If an appropriately selected ballast resistance is present in the electrical circuit, then the CCM grows across the entire width of the insulating gap and rests against the residual oxide. This state corresponds to the limiting curve *4* in Fig. 2. If we switch the ballast resistance out, then the initial branch of the curve in Fig. 3a (up to the maximum) is simply the I–V characteristic of the residual oxide, across which the entire applied voltage falls. Taking its actual thicknesses and its high quality into account, one expects that the main conduction mechanism is tunneling,[10,11] which qualitatively explains the nonlinearity of the curve in the first region. Next, as U increases, the field in the residual oxide intensities sufficient to cause electrical breakdown of the oxide. Its resistance falls abruptly and the current through the structure increases accordingly, ''burning out'' individual segments of the CCM on the anode side and thereby increasing the width of the insulating gap. This latter circumstance is connected with spatial localization of the current pulses (see below), which causes the maximum local power densities in the CCM to occur near the anode. Electrical breakdown of the insulator does not cause irreversible changes in it, since in the first place, the CCM burns out before that on account of its poorer thermal contact with the substrate.[5] The increase in the gap width drastically decreases the field in the residual oxide, which suppresses the state of electrical breakdown and accordingly increases the resistance of the oxide. It is these processes that lead to a decrease in the conductivity of the structure, all the way to its complete switching off, which is what we observe on the descending branch of the I–V characteristic.

This mechanism can also account for the absence of a transition of the structure to the *On* branch over the entire range of voltages U less than the voltage at which the structure switched off. In the case in which the resistance of the residual oxide decreases when it undergoes electrical breakdown, almost all the applied voltage drop is across the CCM. In this case, partial burnout of the latter determines the equilibrium width of the insulating gap corresponding to this value of U. After completion of the process (recovery of the initial resistance of the oxide) the total width of the insulating gap turns out to be increased by the thickness of the residual oxide. Therefore, in order to have sufficient current for switching the structure to the *On* state (repeated forming), which requires a certain field strength in the gap, it is necessary to apply a voltage U_{th} that is somewhat greater than the voltage at which the structure switched off. Thus by into account the effect of the residual oxide on the conduction mechanism of an open MIM sandwich structure with a carbonaceous active medium in a nanometer-size insulating gap, one can explain all the experimentally observed peculiarities of the I–V characteristics.

A fundamental question is, does the electroforming process encompass the entire open perimeter of the sandwich structure, or are there individual formed segments distributed randomly, i.e., more or less uniformly, over the perimeter, or, finally, does all the forming take place only in one such carbonaceous nanostructure? The first possibility is eliminated by a simple estimate: to form the CCM it is necessary to collect a certain dose of electron flux. This dose can be estimated from the condition that the entire perimeter of the insulating gap is the working region in a conventional MIM diode. Taking the actual currents and maximum switching times of the switched-off structure (of the order of seconds) into account in our case gives doses for the case in which the CCM fills the entire perimeter that are several orders of magnitude smaller than needed. The choice between the second and third alternatives is also unambiguous and is based on the following simple experiment. Several topologically unconnected open sandwich structures connected in parallel with one another and in series with a single ballast resistance were subjected simultaneously to electroforming. It always turned out that only one of them was formed while in the others the process did not even begin. This implies that in a single structure the process develops only at one point. This result is easily explained if we take into account the abrupt exponential dependence of the current on the voltage applied to a field emitter, which in our case is some nanoprotrusion of the cathode surface. Since the ballast resistance efficiently limits the growth of the voltage across the structure as the current increases, an automatic selection of a single, most critical element takes place from the set of elements connected in parallel. Taking into account general arguments, namely that for the actual divergence angles of the electron flux emitted from an edge, its thickness l (Fig. 4a), which defines the diameter of the CCM, is of the order of the thickness H of the insulating film, we can say that we are dealing here with a single carbonaceous structure of nanometer size in all three spatial dimensions.

We should call attention to the deep analogy between such an object and a scanning tunneling microscope. In both cases there is a mobile conducting nanotip. Only, in a tunneling microscope its position is controlled by external feedback while in the given case it is controlled by internal feedback within the structure. In the first case the mobility of the tip is a result of mechanical translation, and in the second it is the result of the growth (or recession) of the conducting medium. But in both cases the same mechanism is responsible for the high sensitivity of the current to processes taking place in a nanometer-size insulating gap. In this object the carbonaceous conducting structure, by moving toward the anode, probes itself, as it were. For this reason, measurement of the current–voltage characteristics may be regarded as a method of studying the processes of formation and growth of nanostructures.

Figure 5 displays current–voltage characteristics of an open sandwich structure formed to various states. Electroforming took place in the usual way, but was interrupted at the states corresponding to curves *1–3*, obtained with a ballast resistance of $R_b=5$ MΩ. Then, in the voltage growth phase the corresponding I–V characteristics without the ballast were traced out (curves *1′–3′*), during which the structure switched itself off. After this, it switched itself back on (formed itself) upon application of a somewhat larger voltage and with the ballast resistance present in the circuit. The order in which the curves were obtained corresponds to the numeration in the figure.

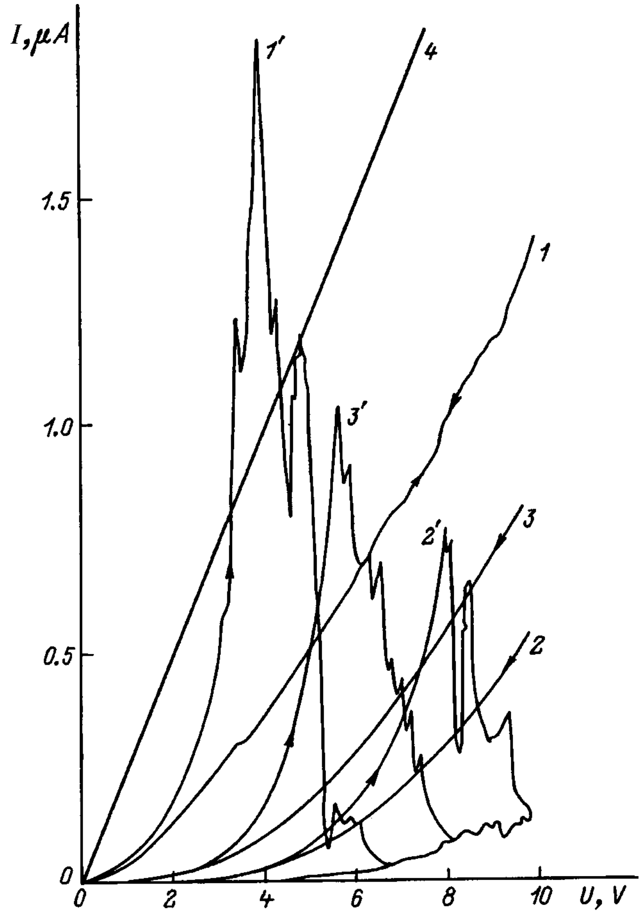

FIG. 5. Current–voltage characteristics of an open MIM sandwich structure with $H=24$ nm.

As is clear from the character of curves *1′–3′*, the corresponding carbonaceous structures differ in more than just the magnitude of the active parameter (or cross section of the CCM) and the corresponding values of the current. Some qualitative parameter is also different from curve to curve, since the positions of the maxima along the U axis are different. This parameter does not vary irreversibly as the processes develop in the structure, rather its value is determined by the state in which the forming cycle was halted (although curves *2′* and *3′* were obtained after *1′*, they correspond to earlier stages within the limits of one cycle). This peculiarity also distinguishes an open sandwich structure from a conventional MIM diode, for which the position of the maximum of the I–V characteristic on the voltage axis (around 4 V), already visible in the first cycle of variation of the voltage, does not subsequently vary, although the amplitude of the current grows significantly.[12] As was noted above, the branch of the curve up to the maximum in our case is the I–V characteristic of the residual oxide, and the maximum corresponds to its electrical breakdown and the burnout of segments of the CCM. It is therefore natural to suppose that the parameter distinguishing the carbonaceous structures corresponding to different curves in Fig. 5 is the thickness of the residual oxide: a thicker oxide layer requires a larger voltage to reach the breakdown field strength. Since the current in an open sandwich structure increases as the structure is formed, i.e., it passes successively through states analogous to *2*, *3*, and *1*, the first segments to form and be connected into the electrical circuit are the segments with thicker residual oxide. But if we look at the actual oblique profile of etching of the anodic oxide [Fig. 4b] in the insulating gap, then this is exactly what should take place as the CCM advances from the cathode to the anode: first the state with residual oxide thickness h_1 is reached, then the state with h_2, etc. (of course, the propagation of the CCM presupposes that at least part of the electron flux goes out into the organic insulator; this process is facilitated by the specific configuration of the electric field in the insulating gap and by the fact that the organic material and aluminum oxide have different dielectric constants, giving rise to an enhanced field strength at the CCM–organic insulator boundary). In the final state, all of them work in parallel; however, since the tunneling current depends exponentially on the width of the insulating gap,[11] the I–V characteristic is governed by the regions with minimum thickness of the insulator. The discontinuous drops in the current at voltages greater than the maximum of the I–V characteristic correspond to burnout of individual regions of the CCM, where the first segments to go are the ones with minimum thickness of the residual oxide, i.e., the lower tips of the carbonaceous structure, since it is specifically in them that the maximum power is dissipated due to the local nature of the current paths through the nanotips of the CCM.

From the proposed mechanism of processes in an open MIM sandwich structure it follows that another interesting effect is possible, namely natural (free) oscillations of the conductivity in such an object. Indeed, suppose we apply a sufficiently large voltage to such a structure, one corresponding to the descending branch of the I–V characteristic. Then, as the CCM grows from the cathode, the insulating gap shrinks and the field strength in it grows. When the field reaches the breakdown value in the residual oxide, the resistance of the latter falls abruptly and almost all of the voltage now falls across the carbonaceous conducting medium. Its lower end burns off to some equilibrium position, which is determined by the magnitude of the applied voltage. The field strength in the oxide falls and the breakdown is ‘‘extinguished.’’ But the state so obtained is not stationary. An applied dc voltage leads to the passage of a significantly smaller, but finite current amplitude that causes the CCM to grow and approach the anode until the next breakdown of the residual oxide, after which the cycle repeats itself, i.e., natural oscillations of the current should be observed in the structure. It is clear that such oscillations will have a stochastic character: there are too many fundamentally random processes and major nonlinearities in the system for it to be otherwise. First of all, these include the atomic discreteness of the material, which can have a large effect, since the dimensions of the insulating gap can be measured in nanometers. In such a situation, the addition of another carbon atom to the CCM can significantly alter the size of the insulating gap and noticeably increase the current through the structure due to its exponential dependence on the size of the gap. Another random factor is the probability of onset of breakdown in the nanometer-size residual oxide and the nature of its development. This factor determines the degree of burnoff of the CCM, which is to say, the width of the insulating gap at which regeneration of the carbonaceous nanostructure begins. In addition, the random noise can be overlaid by dynamical chaos, which is very probable in such a system containing feedback loops and strong nonlinearities.

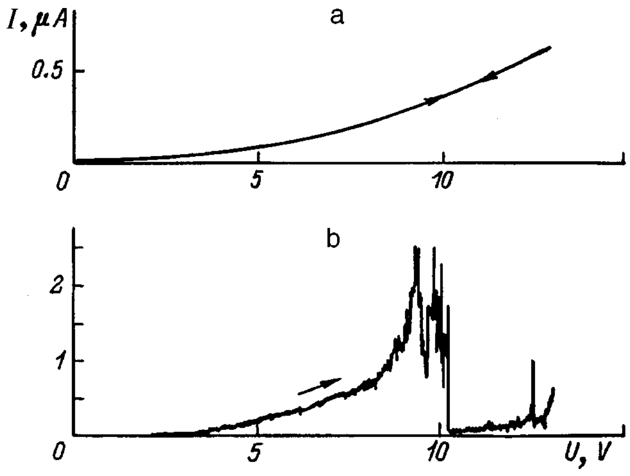

FIG. 6. Current–voltage characteristics of an open MIM sandwich structure with $H=24$ nm: a — limiting forming curve for $R_b=7.5$ MΩ, the thickness of the residual oxide is greater than in Fig. 3; b — voltage sweep rate equal to 0.08 V/s, $R_b=1$ MΩ.

It is advantageous to perform experimental observations of natural oscillations of the current with the ballast resistance included in the circuit, since the latter limits current jumps, thereby decreasing the depth of switching-off of the structure, i.e., it does not permit the appearance of too-large insulating gaps, which facilitates subsequent switching-on in the growth phase of the CCM. Figure 6b shows a I–V characteristic of a structure preformed to the state shown in Fig. 6a. The sweep rate here is an order of magnitude lower than usual. This made it possible to identify in just one pass the range of voltages in which natural oscillations are possible. It can be seen that in the interval from 8 to 10 V one does not observe a simple continuous switching-off of the structure due to burnout of the CCM as in the previous cases, but rather an oscillatory regime arises, with large current surges. Figure 7 displays time plots of the current for structures formed as in Fig. 6a, at a constant voltage taken from the range in which pronounced natural oscillations are present. The varied character of the dependences in these graphs is connected with the different values of the ballast resistance. A large ballast resistance, by efficiently damping the switching-off process, leads to an increase in the mean value of the current and suppresses extinction of the oscillations. For small R_b the natural oscillations that obtain have the form of separate spikes against a background of relatively small currents. As a rule, they are produced by a complete switching-off of the structure to zero current, after which the structure does not switch back on for a fairly extended time.

A study of such processes with high time resolution (down to 2.5 μs) and computer recording of the data showed that a standard recording device strongly smooths out the true time dependence of the current. At large voltages this dependence is a quite smooth curve, against which relatively

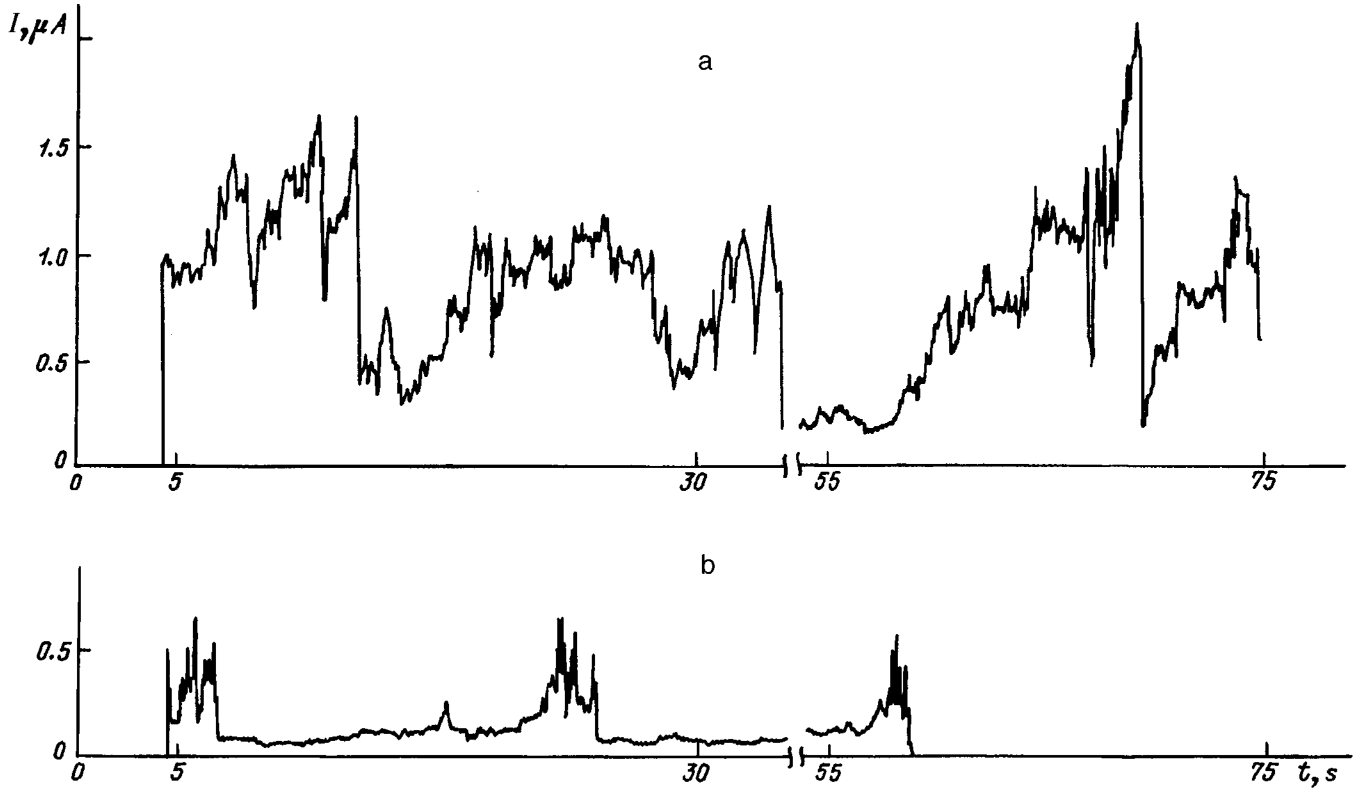

FIG. 7. Time plots of the current of an open MIM sandwich structure for U=8.5 V; R_b=1 (a), 0.3 MΩ (b).

rare chaotic current pulses are observed, having characteristic durations in the range 10–50 μs. The amplitudes of the pulses reach their largest values in (several μA) in the region of the maximum of the I–V characteristic and are significantly lower at large voltages, corresponding to the descending branch. The pulses are absent in the initial segment of the I–V characteristic. These results fit well into the picture of such an object proposed above. The separate pulses can correspond to isolated acts of growth–burnout of the CCM at the molecular level, but the curves in Fig. 7 reflect changes in the geometrical characteristics (envelope) of the carbonaceous nanostructure as a whole. The largest amplitudes of the pulses at voltages near the maximum of the I–V characteristic are due to the fact that it is precisely in this region that the width of the insulating gap is at its minimum, meaning that the relative influence of the appearance (or disappearance) of a particle of the carbonaceous conducting medium in the gap will be at its greatest.

In connection with the above, the question arises as to what extent does the characterization[7] of electroforming as a process of self-organization of a nanometer-size insulating gap in a carbonaceous medium, which is valid for the model of a conventional MIM diode, apply to the object under study here. The presence of a thin insulating layer, rigidly built in and connected in series with the CCM, radically alters the behavior of the structure despite the fact that all of the fundamental physical processes taking place in it remain the same. The fundamental difference is that the appearance of the new element — the insulating layer — changes the attractor of the system. In a conventional MIM diode the final state to which the system tends is a stationary structure (a nanometer-size insulating gap), which corresponds to a static attractor, i.e., it is simply a point in phase space. In the present case, the final state is an oscillatory process with an attractor in the form of a limit cycle or, what is more likely, a strange attractor. Thus, in the present case we must speak not only of spatial, but also temporal self-organization.

In the investigated system, self-organization is manifested in yet another form. We have a self-organizing structure on a nanometer scale not only across the insulating gap, but along it as well. Indeed, the voltage is applied to the entire open perimeter of the sandwich MIM structure, with the result being the formation of a single, isolated carbonaceous nanostructure despite the absence of external organizing factors with such a degree of localization. The reasons for this include the abrupt exponential dependence of the current on the field strength in the insulating gap, on the one hand, and the presence of the ballast resistance in series with the sample during electroforming, on the other. As soon as one nanoprotrusion on the cathode surface begins to emit, part of the voltage falls across the ballast resistance, the voltage on the sandwich structure is decreased, and the carbonaceous conducting medium can no longer form at any other spot.

In view of all we have said, it is more correct to speak of electroforming of such objects as a process of self-organization not of a nanometer-size insulating gap, but more generally of the carbonaceous conducting structure. Such a formulation, being more general, is valid in all cases.

In conclusion, it should be noted that the proposed design of a nano-MIM diode in the form of an open sandwich structure is not only a new and interesting means of studying the phenomenon of electroforming, but also an object very similar to an actually existing nonvolatile memory element based on self-organizing carbonaceous structures.[13]

This work was carried out with the financial support of the State Scientific–Technical Subprogram ‘‘Promising Technologies and Devices of Micro- and Nanoelectronics’’ of the Russian Ministry of Science and Technology.

[1] H. Pagnia and N. Sotnik, Phys. Status Solidi A **108**, 11 (1988).
[2] S. Franzka, G. Knell, H. Pagnia, and N. Sotnik, Int. J. Electron. **76**, 723 (1994).
[3] H. Pagnia, N. Sotnik, and W. Wirth, Int. J. Electron. **69**, 25 (1990).
[4] V. M. Mordvintsev and V. L. Levin, Zh. Tekh. Fiz. **64**(12), 88 (1994) [Tech. Phys. **39**, 1249 (1994)].
[5] V. M. Mordvintsev and V. L. Levin, Zh. Tekh. Fiz. **66**(7), 83 (1996) [Tech. Phys. **41**, 679 (1996)].
[6] V. M. Mordvintsev and V. L. Levin, Zh. Tekh. Fiz. **68**(9), 139 (1998) [Tech. Phys. **43**, 1133 (1998)].
[7] K. A. Valiev, V. L. Levin, and V. M. Mordvintsev, Zh. Tekh. Fiz. **67**(11), 39 (1997) [Tech. Phys. **42**, 1275 (1997)].
[8] V. M. Mordvintsev and V. L. Levin, *Submicron Technology* [in Russian], Tr. FTIAN Vol. 9, Nauka, Moscow (1995), pp. 9–18.
[9] G. A. Vorob’ev and V. M. Gaponenko, Izv. Vuzov. Fiz., No. 1, 65 (1991).
[10] L. L. Odynets and V. M. Orlov, *Anodic Oxide Films* [in Russian] (Nauka, Leningrad, 1990, 200 pp.).
[11] J. G. Simmons, *Tunneling Phenomena in Solids* (Plenum Press, New York, 1969; Mir, Moscow, 1973, pp. 131–142).
[12] H. Pagnia, Int. J. Electron. **69**, No. 1, 33 (1990).
[13] K. A. Valiev, S. E. Kudryavtsev, V. L. Levin *et al.*, Mikroelektronika **26**, No. 1, 3 (1997).

Translated by Paul F. Schippnick

Ultrafast photodetectors based on the interaction of microwave radiation and a photoexcited plasma in semiconductors

V. V. Antonov, S. V. Ivanov, V. P. Tsarev, and V. N. Chupis

Saratov State Technical University, 410034 Saratov, Russia

(Submitted February 11, 1997)
Zh. Tekh. Fiz. **68**, 94–98 (November 1998)

The interaction of microwave radiation with the plasma in photoionized semiconductor photocells (CdS, CdSe) placed in waveguide measurement systems is investigated theoretically and experimentally. The interaction of the characteristic waveguide modes with a photoexcited semiconductor plasma is investigated. The dependence of the reflection coefficient and phase of the microwave radiation on the intensity of the optical radiation to be measured is obtained, and the influence of the surface of the semiconductor photocells on these parameters is investigated. A microwave photodetector design based on a millimeter-wave interferometer is developed. © *1998 American Institute of Physics.* [S1063-7842(98)01811-X]

One of the more important directions of development of semiconductor photometry has to do with the need to vary the parameters of short and ultrashort (as low as 1 ns) pulses of optical (laser) radiation. The creation of measurement systems of this kind is of fundamental importance, both to provide a metric for a new class of high-power pulsed sources of coherent optical radiation and to investigate the properties of various materials under the action of laser pulses.

However, the traditional principles of semiconductor photometry, based on measuring the parameters of a photoexcited semiconductor at constant current, have finally exhausted practically all their possibilities. The main and fundamentally unavoidable limitation on the response time of such traditional photodetectors as photodiodes and photoresistors derives from the low mobility of ionized carriers. The creation of large extracting fields in the given case does not solve the problem, since generation–recombination noise and heating of the semiconductor element are greatly increased under these conditions. Another characteristic drawback of this measurement principle has to do with limitation of the sensitivity of semiconductor photodetectors by recombination processes, as a consequence of which not all the electrons excited into the conduction band make it from the ‘‘cathode’’ to the ‘‘anode.’’ Rather, a significant fraction of them recombine, which naturally lowers the sensitivity and accuracy of the method.

One of the most promising directions of research in this field consists in using the interaction of weak microwave radiation with a photoexcited (photoionized) plasma in a semiconductor. The primary advantage and distinguishing feature of this effect is that weak microwave radiation ‘‘registers’’ the carriers at the instant they are excited into the conduction band. The current level of development of microwave technology in the millimeter and submillimeter bands makes it possible to carry out such measurements in the frequency range 37–150 GHz during times of 10 ps or less.

The most promising measurement system is a microwave interferometer which has semiconductor inserts (photodetector elements) as its reflecting mirrors. In such a case, as theoretical and experimental studies have shown, the change of the phase of the wave reflected from the semiconductor is directly proportional to the concentration of photoexcited carriers.

The bulk character of the interaction of the microwave field with the semiconductor and the presence of resonance effects in the given frequency range substantially expand the possibilities for the development of semiconductor photometry for a wide class of materials. To determine the main parameters of microwave photodetectors, it is necessary to examine the peculiarities of the interaction of microwave radiation with a photoionized semiconductor in a channeling electrodynamic system (e.g., in the aforementioned waveguide system of a microwave interferometer).

Let us consider the interaction of an H_{10} electromagnetic wave propagating in a rectangular waveguide with a photoexcited semiconductor slab which completely covers the cross section of the waveguide system and is nonuniform in the direction of propagation z. To measure the parameters of optical radiation, one customarily uses thin semiconductors and semiconductor films with thickness substantially less than the characteristic diffusion length. In this case, it is necessary to take into account the strong influence of the surface on the distribution and concentration of the photoexcited electrons. In such photodetector elements the carriers can diffuse to the surface and recombine there. The concentration distribution under such conditions is defined by the ratio L_s/L, where $L_s=D/P_s$, with P_s the surface recombination rate and D the ambipolar diffusion coefficient.

Let us consider a steady-state distribution of electrons and holes in a semiconductor irradiated by light with intensity I_0. The intensity of the optical radiation I inside the semiconductor layer is given by the well-known relation[1,2]

$$I=I_0(1-R_s)\exp(-\alpha\cdot z), \tag{1}$$

where R_s is the reflection coefficient of light from the surface $z=0$, and α is the attenuation coefficient.

The total current density in the sample is determined by the electron and hole diffusion currents

$$j_n = en\mu_n E_z + eD_n \frac{dn}{dz}, \tag{2}$$

$$j_p = en\mu_p E_z - eD_p \frac{dp}{dz}, \tag{3}$$

$$\mathbf{j} = \mathbf{j}_n + \mathbf{j}_p, \tag{4}$$

where

$$D_n = \frac{\mu_n kT}{e}, \quad D_p = \frac{\mu_p kT}{e}$$

are the diffusion coefficients of the electrons and holes, and E_z is the intensity of the associated electric field.

We represent the equations of continuity in standard form[2]

$$\frac{\partial(\delta n)}{\partial t} = q_n - \frac{\delta n}{\tau_{n\ \mathrm{rec}}} + \frac{1}{e}\frac{dj_n}{dz} = 0, \tag{5}$$

$$\frac{\partial(\delta p)}{dt} = q_p - \frac{\delta p}{\tau_{p\ \mathrm{rec}}} - \frac{1}{e}\frac{dj_p}{dz} = 0, \tag{6}$$

where $\delta n = n - n_0$ and $\delta p = p - p_0$, $\tau_{n\ \mathrm{rec}}$ and $\tau_{p\ \mathrm{rec}}$ are the recombination times of the electrons and holes, q_n and q_p are the generation rates of the electrons and holes, and n_0 and p_0 are the equilibrium concentrations of the electrons and holes.

It follows from system (2)–(4) that

$$E_z = \frac{D_p \dfrac{dp}{dz} - D_n \dfrac{dn}{dz}}{n\mu_n + p\mu_p}. \tag{7}$$

As a result of the action of the field E_z, a space charge ρ exists during the Maxwellian relaxation time τ_m (Ref. 2)

$$\rho(z) = \rho(z,0)\exp(-t/\tau_m). \tag{8}$$

For a number of semiconductors the time τ_m is much less than the lifetimes $\tau_{n\ \mathrm{rec}}$ and $\tau_{p\ \mathrm{rec}}$; therefore it may be assumed that $\rho(z) \approx 0$ (Ref. 1) and, correspondingly, $\delta n \approx \delta p$. Substituting $\delta n \approx \delta p$ $(n \approx p)$ into Eq. (7), we obtain the expression

$$E_z = \frac{D_p - D_n}{n\cdot(\mu_n + \mu_p)}\frac{d(\delta n)}{dz},$$

which together with Eqs. (2) and (5) gives the following equation for the electron concentration δn:

$$\frac{\partial(\delta n)}{\partial t} = D\frac{d^2\delta n}{dz^2} + q_0\exp(-\alpha z) - \frac{\delta n}{\tau_{n\ \mathrm{rec}}}, \tag{9}$$

where

$$q_0 = \frac{I_0(1-R_s)}{\hbar\omega},$$

and $\hbar\omega$ is the energy of one of the photons incident on the semiconductor slab.

The general solution of Eq. (9) has the form[1]

$$\delta n = \frac{q_0\tau_{n\ \mathrm{rec}}}{1-\alpha^2L^2}\exp(-\alpha z) + C_1\exp(-z/L) + C_2\exp(z/L), \tag{10}$$

where $L = \sqrt{D\cdot\tau_{n\ \mathrm{rec}}}$ is the ambipolar diffusion length.

We determine the integration constants C_1 and C_2 from the boundary conditions[1]

$$j_n(z=0) = P_{s1}\cdot\delta n(z=0),$$

$$j_n(z=d) = P_{s2}\cdot\delta n(z=d), \tag{11}$$

where $|P_{s1}| = |P_{s2}| = P_s$ are the surface recombination rates at the faces $z=0$ and d; $P_{s1} > 0$ and $P_{s2} > 0$ if the current is directed into the interior of the sample from its boundaries.[1]

In terms of the customary notation

$$a_0 = \frac{D}{L} - P_s, \quad a_1 = \frac{D}{L} + P_s, \quad b_0 = \frac{q_0\tau_{n\ \mathrm{rec}}}{1-\alpha^2L^2}$$

system (11) admits a solution of the form

$$C_1 = \frac{b_0[\alpha D\cdot(a_1\exp(d/L) - a_0\exp(-\alpha d)) + P_s(a_1\exp(d/L) + a_0\exp(-\alpha d))]}{a_0^2\exp(-d/L) - a_1^2\exp(d/L)},$$

$$C_2 = \frac{b_0[\alpha D\cdot(a_0\exp(-d/L) - a_1\exp(-\alpha d)) + P_s(a_0\exp(-d/L) + a_1\exp(-\alpha d))]}{a_0^2\exp(-d/L) - a_1^2\exp(d/L)}.$$

An H_{10} electromagnetic wave, incident on such a slab, is characterized by the following field structure inside the waveguide:

$$E_x^0 = 0, \quad E_y^0 = -\frac{j\omega\mu_0}{\pi}\alpha H_0\sin\left(\frac{\pi x}{a}\right)\exp(-j\beta z),$$

$$E_z^0 = 0,$$

$$H_x^0 = \frac{j\beta a}{\pi}H_0\sin\left(\frac{\pi x}{a}\right)\exp(-j\beta z), \quad H_y^0 = 0,$$

$$H_z^0 = H_0\cos\left(\frac{\pi x}{a}\right)\exp(-j\beta z), \tag{12}$$

where a is the length along the wide wall of the waveguide, which is aligned with the x axis;

$$\beta=\sqrt{\left(\frac{\omega}{c}\right)^2-\left(\frac{\pi}{a}\right)^2},$$

H_0 is the amplitude of the magnetic component of the wave, ω is the frequency of the field, and μ_0 is the magnetic permeability of free space.

The z axis points in the direction of propagation of the wave. The wave reflected from the surface of the semiconductor (at $z=0$) has the following components:

$$E_x^1=0,\ E_y^1=-\frac{j\omega\mu_0}{\pi}\alpha H_0^1\sin\left(\frac{\pi x}{a}\right)\exp(i\beta z),\ \ E_z^1=0,$$

$$H_x^1=-\frac{j\beta a}{\pi}H_0^1\sin\left(\frac{\pi x}{a}\right)\exp(j\beta z),\ \ H_y^1=0,$$

$$H_z^1=H_0^1\cos\left(\frac{\pi x}{a}\right)\exp(j\beta z). \tag{13}$$

In the region of the waveguide filled by the semiconductor, the electric and magnetic components of the field are determined by the system of equations

$$E_y=A(z)\sin\frac{\pi}{a}x,\ \ H_x=\frac{\partial E_y}{\partial z}\frac{1}{j\omega\mu_0},$$

$$H_z=-\frac{\partial E_y}{\partial x}\frac{1}{j\omega\mu_0}.$$

The function $A(z)$ in the given case satisfies the differential equation

$$\frac{\partial^2 A(z)}{\partial z^2}+\frac{\omega^2}{c^2}\left(\varepsilon_r+\frac{\sigma}{j\omega\varepsilon_0}-\left(\frac{\pi}{a}\frac{c}{\omega}\right)^2\right)\cdot A(z)=0, \tag{14}$$

where ε_r is the dielectric constant of the semiconductor lattice, and

$$\sigma=c^2(\delta n+n_0)\cdot\left(\frac{1}{m_n(j\omega+\nu_n)}+\frac{1}{m_p(j\omega+\nu_p)}\right).$$

If the dimension of the slab along the axis of propagation of the wave is chosen to be less than α^{-1} and L, then the excess concentration $\delta n(z)\approx\delta n(0)\approx\text{const}$ (Ref. 1), so that Eq. (14) can be solved by the Wenzel–Kramers–Brillouin (WKB) method.[3] Defining the function $A(z)$ in the form

$$A(z)=\left[A_0(z)+\frac{c}{\omega}A_1(z)+\frac{c^2}{\omega^2}A_2(z)+\ \dots\ +\frac{c^n}{\omega^n}A_n(z)\right]$$
$$\times\exp\left(-j\frac{\omega}{c}\psi(z)\right),$$

we obtain the following solution:[3]

$$A(z)=\varepsilon^{-1/4}\cdot\left(C_1\exp\left(-j\frac{\omega}{c}\int_0^z\sqrt{\varepsilon}dz\right)\right.$$
$$\left.+C_2\exp\left(j\frac{\omega}{c}\int_0^z\sqrt{\varepsilon}dz\right)\right), \tag{15}$$

where

$$\varepsilon(z)=\varepsilon_r+\frac{\sigma}{j\omega\varepsilon_0}-\left(\frac{\pi}{a}\frac{c}{\omega}\right)^2.$$

The solution of Eq. (14) in the form (15) can be obtained provided the inequality

$$\frac{c}{\omega}\left|\frac{A_1(z)}{A_0(z)}\right|\ll 1.$$

is satisfied.

For the semiconductors we used in our experiments (CdSe and CdS) the indicated ratio varies within the limits

$$0.03<\frac{c}{\omega}\left|\frac{A_1(z)}{A_0(z)}\right|\approx\frac{1}{2\sqrt{2}}\frac{1}{L\sqrt{\varepsilon_r}}\frac{c}{\omega}<0.1,$$

so that a solution of the form (15) satisfies Eq. (14) to a good approximation.

The estimates obtained allow us to determine the y component of the electric field **E** and the x component of the magnetic field **H**:

$$E_y=[C_1\cdot F_1(z)+C_2\cdot F_2(z)]\sin\left(\frac{\pi}{a}x\right)\cdot\varepsilon^{-1/4},$$

$$H_x=[C_1\cdot f_1(z)+C_2\cdot f_2(z)]\sin\left(\frac{\pi}{a}x\right), \tag{16}$$

where

$$F_1(z)=\exp\left(-j\frac{\omega}{c}\int_0^z\sqrt{\varepsilon}dz\right),$$

$$F_2(z)=\exp\left(j\frac{\omega}{c}\int_0^z\sqrt{\varepsilon}dz\right),$$

$$f_1(z)=-\frac{1}{j\omega\mu_0}\left(\frac{1}{4}\varepsilon^{-5/4}\frac{\partial\varepsilon}{\partial z}+\varepsilon^{1/4}j\frac{\omega}{c}\right)\cdot F_1(z),$$

$$f_2(z)=\frac{1}{j\omega\mu_0}\left(-\frac{1}{4}\varepsilon^{-5/4}\frac{\partial\varepsilon}{\partial z}+\varepsilon^{1/4}j\frac{\omega}{c}\right)\cdot F_2(z).$$

Equating the tangential components of the field at the boundaries $z=0$ and d, we obtain an algebraic system of equations in the reflection coefficients $R=H_0^1/H_0$, $R_1=C_2/C_1$ at $z=0$ and $z=d$, respectively:

$$\frac{1+R}{j\beta(R-1)}=\frac{1}{\varepsilon^{1/4}(0)}\left(\frac{1+R_1}{f_1(0)+R_1 f_2(0)}\right),$$

$$-\frac{1}{j\beta}=\frac{1}{\varepsilon^{1/4}(d)}\left(\frac{F_1(d)+R_1F_2(d)}{f(d)+R_1 f_2(d)}\right). \tag{17}$$

The system of equations (17) has the solution

$$R=\frac{1+j\beta F_3}{j\beta F_3-1},\ \ R_1=\frac{j\beta F_1-\varepsilon^{1/4}(d)\cdot f_1(d)}{j\beta F_2+\varepsilon^{1/4}(d)\cdot f_2(d)},$$

where

$$F_3=\frac{1}{\varepsilon^{1/4}(0)}\left(\frac{1+R_1}{f_1(0)+R_1 f_2(0)}\right).$$

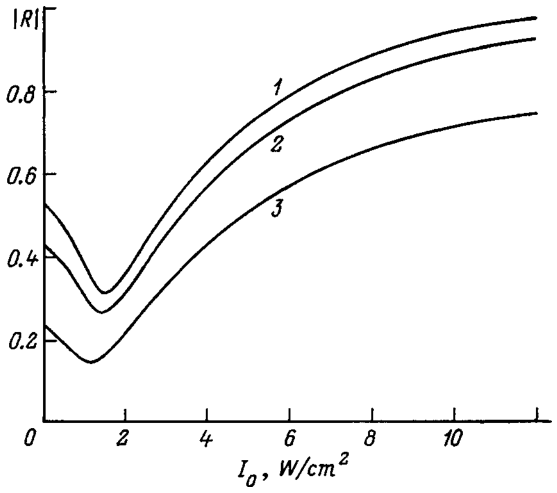

FIG. 1. Dependence of the modulus of the reflection coefficient on the the light intensity for $P_s = 10$ m/s; $d = 300$ (*1*), 200 (*2*), 100 μm (*3*).

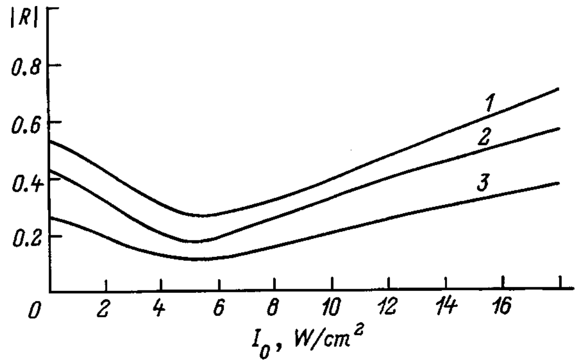

FIG. 3. Dependence of the modulus of the reflection coefficient on the light intensity for for $P_s = 50$ m/s (*1–3* — the same as in Fig. 1).

Figure 1 plots the modulus of the reflection coefficient of microwave radiation as a function of the light intensity I_0 at the microwave frequency $f = 37.5$ GHz for different values of the thickness of a CdS slab in the direction of propagation of the microwave radiation. The curves of this dependence are nonmonotonic, and the minima of $|R|$ shift to higher intensities as d is increased. Increasing the intensity causes charge carriers to appear in the volume of the semiconductor and results in a decrease in the real part of ε, which leads to matching of the media at the vacuum–semiconductor interface. As the thickness of the sample is decreased, the amplitude of the wave reflected from the second face grows, and at the boundary $z = 0$ the two oscillations, phase-shifted relative to each other, add. As a result of this superposition of waves, the modulus of the reflection coefficient decreases with decrease of d. The surface recombination rate P_s in the given case is equal to 10 m/s.

Figure 2 plots quantitative estimates of the phase of the wave reflected from the $z = 0$ face as a function of the light intensity. In the interval of values from 0 to 1.5 W/cm^2 the dependence of the phase on the intensity I_0 is almost linear.

The dependence of the modulus of the reflection coefficient $|R|$ and its phase on the intensity of the optical radiation for $P_s = 50$ m/s is plotted in Figs. 3 and 4. The decrease in the charge carrier concentration associated with surface recombination of carriers leads to a more gradual dependence $R(I)$. Matching of the media for Re $\varepsilon = 0$ is observed for $P_s = 50$ m/s at a higher light intensity than in the case $P_s = 10$ m/s. As can be seen, in the intensity interval from

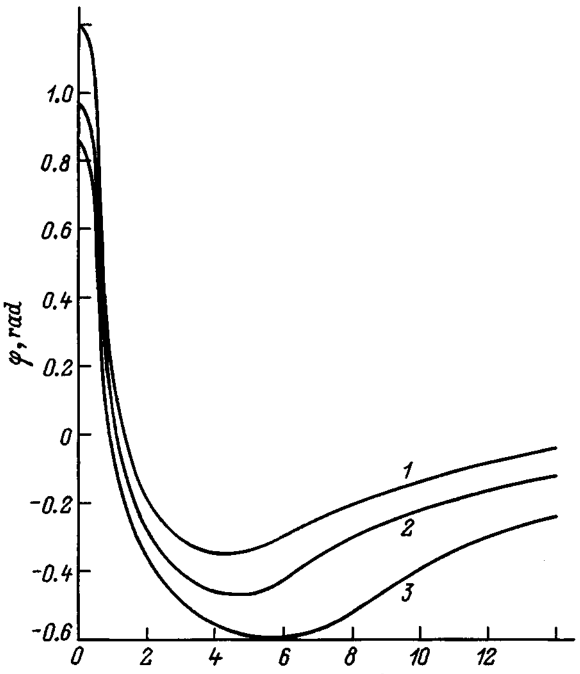

FIG. 2. Dependence of the phase of the reflection coefficient on the light intensity for $P_s = 10$ m/s (*1–3* — the same as in Fig. 1).

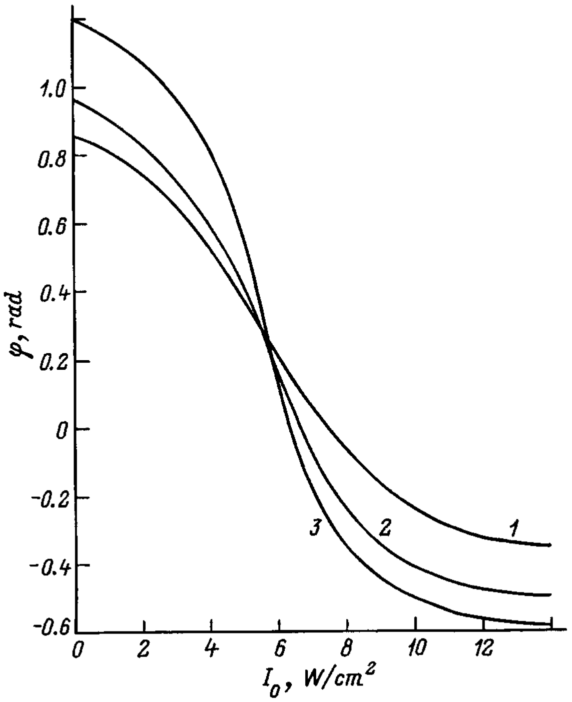

FIG. 4. Dependence of the phase of the reflection coefficient on the light intensity for $P_s = 50$ m/s (*1–3* — the same as in Fig. 1).

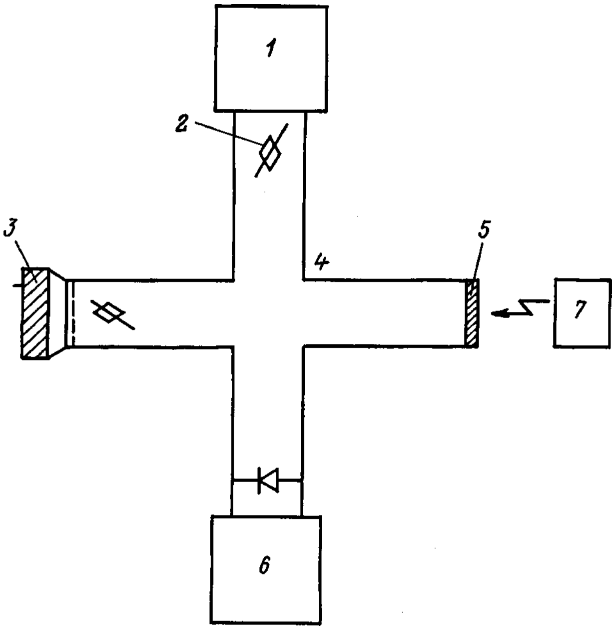

FIG. 5. Block diagram of a microwave interferometer: *1* — GCh-156 generator, *2* — attenuator, *3* — shorting plunger, *4* — waveguide double tee, *5* — semiconductor element, *6* — microwave detector with V6-4 microvoltmeter, *7* — laser.

1.0 to 8 W/cm^2 the dependence $\varphi(I)$ is described by a linear function.

The experimental system in the given case (Fig. 5) consists of a microwave interferometer based on a waveguide double tee *4* in circuit with a generator *1* for the eight-millimeter band. The arms of the interferometer are closed off by the CdS semiconductor element *5*, an attenuator *2*, and a shorting plunger *3*. The principle of operation of the microwave interferometer is based on contact-free measurement of the parameters of the photoexcited semiconductor element. Waves reflected from the semiconductor and the shorting plunger interfere in the symmetry plane of the waveguide tee.[5] If the maximum of the electric field is located in this plane, then energy does not enter the arm with the detector *6*. When the semiconductor element is acted on by optical radiation from an LG-38 He–Ne laser with wavelength $\lambda=0.63$ μm, the charge carrier concentration in it varies. This leads to a change in the reflection coefficient and the appearance of a signal in the indicator circuit. During the measurement of the dependence $R(I)$ the intensity of the laser radiation is regulated by a polarizer.

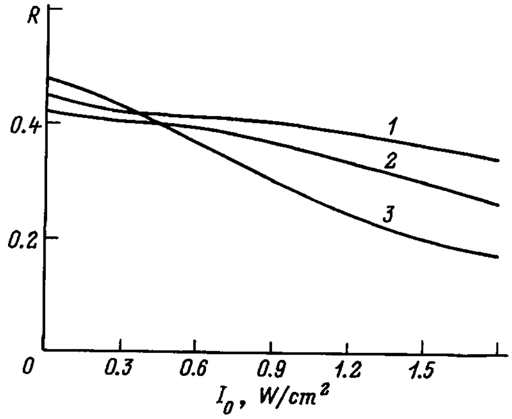

FIG. 6. Experimental curves of the dependence of the modulus of the reflection coefficient on the laser intensity I_0 for $f=27$ (*1*), 36 (*2*), and 29.8 GHz (*3*).

Figure 6 plots experimental curves of the dependence $R(I)$ for sample thickness $d=300$ μm and cross section 7.2 $\times$3.4 mm for three microwave frequencies. As can be seen from the figure, the experimental curves are in qualitatively good agreement with the theoretical results.

To summarize, the principle of contact-free measurement of the parameters of a photoexcited plasma by means of microwave radiation makes it possible to broaden the possibilities of semiconductor photometry in the direction of creating ultrafast photodetectors with high accuracy of measurement of the optical radiation parameters.

[1] B. T. Boĭko and Yu. G. Gurevich, *Physics of Solar Cells* [in Russian] (Kharkov State Univ. Press, Kharkov, 1992).
[2] S. P. Kireev, *Physics of Semiconductors* [in Russian] (Vysshaya Shkola, Moscow, 1975).
[3] V. L. Ginzburg, *Propagation of Electromagnetic Waves in Plasma* [in Russian] (Nauka, Moscow, 1967).
[4] L. I. Kats, V. P. Tsarev, and V. N. Chupis, Inventor's Certificate No. 1185259 (USSR); International Classification OIP 21/12, No. 375621/24/09; publ. Byull. Izobret., No. 38 (1985).
[5] L. A. Dushin, *Microwave Interferometers for Measuring the Density of a Plasma in a Pulsed Gas Discharge* [in Russian] (Atomizdat, Moscow, 1973).
[6] V. N. Chupis and V. P. Tsarev, in *Defense Technology* [in Russian], Scientific–Technical Collection (Moscow, 1996), 35 pp.

Translated by Paul F. Schippnick

X-ray converters for radiation treatment of thin films

V. I. Bespalov

Tomsk Polytechnical Institute, 634004 Tomsk, Russia

V. V. Ryzhov and I. Yu. Turchanovskiĭ

Institute of High-Current Electronics, Siberian Branch of the Russian Academy of Sciences, 634055 Tomsk, Russia

(Submitted June 30, 1997)
Zh. Tekh. Fiz. **68**, 99–101 (November 1998)

The energy absorbed in thin films of selected materials bombarded by x rays emitted in the braking of low-energy electrons ($E_0 < 500$ keV) in converters with various atomic numbers ($Z = 29 - 73$) is calculated by the Monte Carlo method. The program takes into account both of the *K*-shell ionization mechanisms that lead to emission of characteristic photons as a result of electron impact and as a result of the photoelectric effect, and the characteristic radiation is shown to make a large contribution to the absorbed energy in thin films. Calculations show that the proper choice of material and thickness of the converter affords a two- to fivefold increase in the energy of the x radiation absorbed in thin films of semiconductor materials.

[S1063-7842(98)01911-4]

INTRODUCTION

Trends in the generation of bremsstrahlung arising in the braking of high-energy electrons in a target have been well examined both theoretically and experimentally.[1,2] As a rule, materials with a high atomic number Z (tungsten and tantalum) are used to obtain a maximal bremsstrahlung yield at these energies. However, in the braking of low-energy electrons, a considerable fraction of the energy can be emitted in the form of the characteristic radiation that arises both as a result of photoabsorption of secondary photons and as a result of ionization of the inner shells of the atom by electron impact. Since the cross section of the secondary process grows as the atomic number Z is decreased, a high efficiency of generation of x radiation is also possible in materials with low Z. This circumstance must be taken into account in the radiation treatment of thin films and foils, where the efficiency is determined not only by the total energy of the radiation but also by its spectral composition.

PROGRAM

With the aim of investigating the optimal conditions of irradiation and the choice of efficient x-ray converters for treating thin films, we used a program that performs a Monte Carlo simulation of the development of an electron–photon cascade in a converter and the absorption of the radiation in the film material. For the calculations of the spectral distribution of the radiation on the other side of the converter, a modification of the program was developed[3] that took into account processes of generation of the characteristic radiation both by secondary bremsstrahlung and by electrons. The *K*-shell ionization cross section by electron impact was taken from Ref. 4.

Note that in most Monte Carlo programs simulating the passage of electrons and photons through matter, the characteristic radiation is either not taken into account or is taken into account only through the inner-shell channel of ionization due to the photoelectric effect. This is because the characteristic radiation does not play a substantial role in traditional problems associated with the calculation of the total energy of bremsstrahlung and radiation protection.

CALCULATED RESULTS

To investigate the optimal conditions of radiation treatment of thin films, we calculated the absorbed energy in films of gold and semiconductor materials for converters with the representative atomic numbers $Z = 29$, 42, 57, 73. The initial electron energy was varied from 100 to 500 keV. The thickness of the converters was optimized with respect to the total yield of radiation energy, and a graphite block of appropriate thickness was placed behind the converter to absorb electrons escaping from it.

The important role of characteristic radiation in problems connected with energy absorption in a calorimeter based on a thin gold foil ($d = 27.9$ μm) was demonstrated in Ref. 1. One unexpected result obtained by the authors of this work is that for lanthanum converters ($Z = 57$) roughly 20% more energy is absorbed in the calorimeter than for a titanium converter ($Z = 73$). At the same time, the total bremsstrahlung yield for a tantalum converter should be roughly that much higher.

The calorimeter used in the experiments measured the energy absorbed in a gold foil, whose thickness was insufficient for total absorption of all the radiation energy. Therefore we investigated the influence of the foil thickness on the amount of energy absorbed in the foil (Fig. 1). It follows from Fig. 1 that the readings of the calorimeter with a thin gold foil behind a lanthanum converter are higher than for a tantalum converter; however, at greater thicknesses of

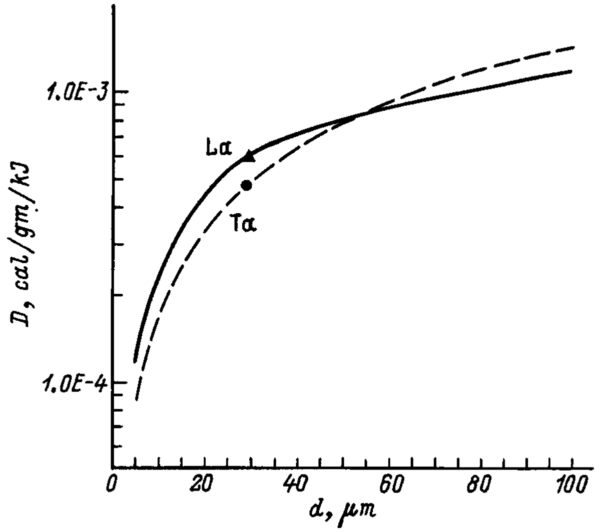

FIG. 1. Energy absorbed in the calorimeter versus thickness of the gold foil for electrons with initial energy $E_0 = 1.0$ MeV incident on a lanthanum (solid curve) or tantalum (dashed curve) converter of optimal thickness.

the absorbing foil (according to the calculations, for $d > 50$ μm) the situation corresponds to the usual picture of the dependence of the bremsstrahlung yield on Z.

TABLE I. Energy (keV) of the photons exiting a converter of optimal thickness, and fraction (%) of the characteristic emission in it (results of calculations normalized to one incident electron).

E_0, keV	Cu	Mo	La	Ta
50	0.098	0.0608	0.0516	0.060
	69	40	16	0
100	0.235	0.224	0.221	2.43
	54	34	16	5
500	2.26	3.10	4.08	5.08
	5	8	9	8

An analysis of the results of our calculations supported the conclusion of the authors of Ref. 1 that the reason for this effect is the high contribution of the characteristic radiation to the energy absorbed in the thin gold foil. Our values of thecontribution of the characteristic radiation to the energy absorbed in the calorimeter are plotted in Fig. 2. It can be seen from Fig. 2 that the contribution of the characteristic radiation depends on the converter material and the thickness of the calorimeter foil. For thin gold foils ($d < 30$ μm for lanthanum converters and $d < 100$ μm for tantalum converters) the fraction of the energy of the characteristic radiation absorbed in the calorimeter grows and reaches 45% and 20% for lanthanum and tantalum converters, respectively. It should be noted that the contribution of the characteristic radiation formed by electron ionization of the K shell reaches 50% for lanthanum relative to the total energy of the bremsstrahlung and is essentially independent of the foil thickness, while for the tantalum converters it does not exceed 20%. This is explained by the corresponding dependence of the K-shell ionization cross section on Z. The important role of the characteristic radiation in the radiation spectrum for converters with varying atomic number Z, optimized for the total radiation energy, follows from the data shown in Table I. The calculations were performed for a planar geometry: a unidirectional, monoenergetic electron beam is incident on a converter (Cu, Mo, La, Ti), behind which is an electron absorber (a carbon block of appropriate thickness) and, behind it, a film of the material being irradiated (Si, Ge, Au). It follows from the table that up to initial electron energies $E_0 = 100$ keV the total energy yield of the x radiation behind the copper converters is higher than that behind the tantalum converters. The contribution of the characteristic radiation for copper exceeds 50%, whereas for tantalum it is an order of magnitude lower.

Results of calculations for silicon films irradiated by x rays from electrons with initial energy 500 keV are plotted in Fig. 3. The calculations show that the choice of the converter material depends on both the thickness of the foil being

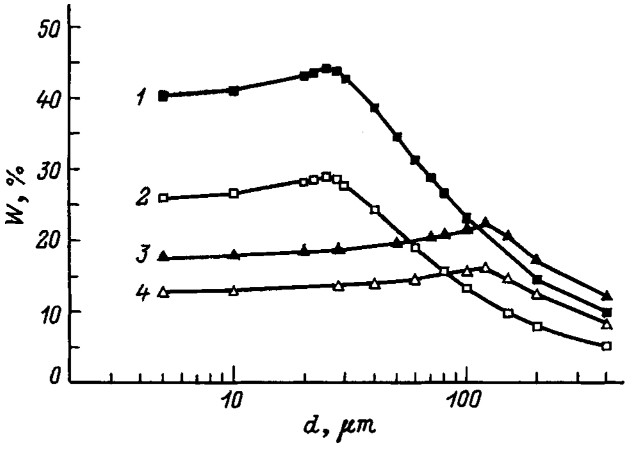

FIG. 2. Contribution of the characteristic K-shell radiation to the energy absorbed in the calorimeter, plotted as a function of the thickness of the gold foil: *1, 3* — calculated with ionization of the K shells by electrons and bremsstrahlung taken into account; *2, 4* — calculated with only ionization of the K shells by bremsstrahlung taken into account.

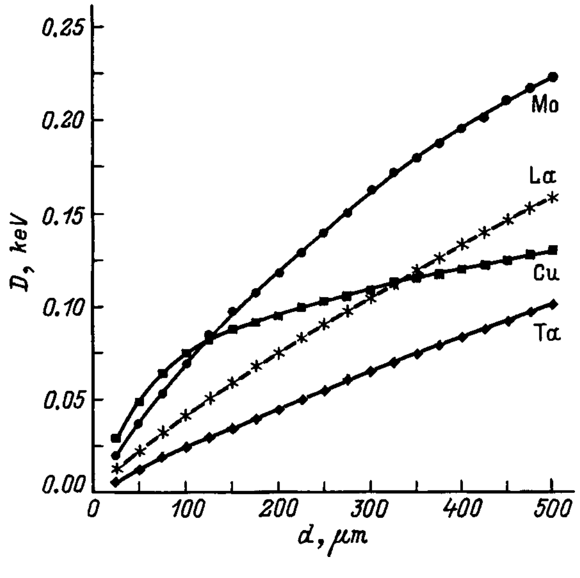

FIG. 3. Dependence of the radiation energy absorbed in the silicon layer on its thickness and the converter material for initial electron energy $E_0 = 500$ keV.

treated and the initial electron energy. Thus, for $E_0 = 500$ keV when treating thin films of silicon ($d < 100$ μm) it is necessary to use copper converters, whereas for thick films ($100 < d < 500$ μm) molybdenum converters are the most efficient. Notice that the energy absorbed in the silicon film (see Fig. 3) is two times higher than in a film of the conventional converter material tantalum. For lower initial electron energies the most efficient converter material is copper, for which there is a 5–6-fold an improvement in terms of the absorbed energy at $E_0 = 100$ keV as compared to tantalum.

Calculations showed that when treating germanium, molybdenum converters should be used only for very thin films ($d < 20$ μm), while for films of medium thickness ($20 < d < 200$ μm) lanthanum converters are the most efficient.

CONCLUSIONS

When an electron beam brakes in a target, the energy of the characteristic radiation arising as a result of *K*-shell ionization in low-atomic-number converters can exceed half the total energy of the x radiation, which gives us a basis to recommend them as the most efficient converters for radiation treatment of thin films and coatings.

A correct Monte Carlo calculation of the generation of x radiation in the braking of low-energy electrons is possible only if one takes into account ionization of inner shells not only as a result of the photoelectric effect but also as a result of electron impact. Basing the choice of the material and thickness of the converter on such calculations makes it possible to increase the energy of the x radiation absorbed in thin films of semiconductor materials by two- to fivefold.

[1] J. A. Halbleib, G. J. Lockwood, and G. H. Miller, IEEE Trans. Nucl. Sci. **NS-23**, 1881 (1976).

[2] V. V. Ryzhov and A. A. Sapozhnikov, in *Proceedings of the Ninth International Conference on High-Power Particle Beams ''Beams-92,''* Washington (1992), Vol. 2, pp. 1199–1204.

[3] V. I. Bespalov, S. D. Korovin, V. V. Ryzhov, and I. Yu. Turchanovsky, in *Proceedings of the Tenth IEEE International Pulsed Power Conference*, Albuquerque, New Mexico (1995), Vol. 1, pp. 75–79.

[4] H. Kolbenstvedt, J. Appl. Phys. **38**, 4785 (1967).

Translated by Paul F. Schippnick

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Color coding of images of deformed zones of diffusely scattering surfaces during optical processing of photographs of projected fringes

A. M. Lyalikov

Ya. Kupala Grodno State University, 230023 Grodno, Belarus

(Submitted August 19, 1997)
Zh. Tekh. Fiz. **68**, 102–105 (November 1998)

An optical method is developed wherein different colors are imparted to zones of the image of a diffusely reflecting flat surface corresponding to different degrees of deformation of the surface. The method is based on principles of spatial filtering during optical processing of photographs of projected fringes in white light. Experimental verification of the method is presented. © *1998 American Institute of Physics.* [S1063-7842(98)02011-X]

In studies of flat surfaces, primary information about deformations and surface stresses is contained in the first derivatives $\partial\omega/\partial x$ and $\partial\omega/\partial y$ of the normal displacements of the surface $\omega(x,y)$, where the x and y axes lie in the plane of the surface.[1] The simplest method of differentiating such displacement data is optical differentiation of interferograms or moiré patterns.[1–4] In studies of comparatively large surface displacements, there has been a movement away from methods of holographic interferometry to less sensitive methods of projection of fringes or deposition of grids on the investigated surface.[1,2,5,6] The application of optical methods for processing photographs of projected fringes has made it possible to adjust the contour bands in an arbitrary way and to control the sensitivity of the measurements of the relief of the investigated surface.[7–9]

The present paper considers a further refinement of the method developed in Ref. 10, which allows one to visualize deformed zones of diffusely scattering surfaces for optical processing of photographs of projected fringes using spatial filtering. It is demonstrated that one can arrange for the images of zones of a diffusely scattering flat surface having different degrees of deformation to be colored in different hues. The given effect is achieved by optical processing of photographs of projected fringes in white light.

Color coding of information about the parameters of the state of an object has been used previously in the optical processing of holograms of a phase object for visualization of the angles of deflection of light rays in the object,[11,12] and also to determine the direction of the gradient of the refractive index.[13,14] Pseudocolor coding has been used for recognition of various brick walls, for example.[15] In connection with a study of surface relief it has been proposed to color code surface regions having the same depth by projecting a grid in white light onto the investigated surface.[16] The color coding in the given method is based on the Talbot effect, which arises when a periodic structure is illuminated in white light.

Let us consider aspects of white-light optical processing of photographs of projected fringes in the study of deformations of a flat, diffusely scattering surface. The amplitude transmittance of such a photograph of projected fringes is equal to[8,10]

$$\tau(x,y)\sim\left\{1+\cos\left[\frac{2\pi y}{T}+\Phi(x,y)\right]\right\}^{-\gamma/2}, \tag{1}$$

where T is the period of the observed fringes in an undeformed region of the flat surface, and γ is the contrast coefficient of the photographic emulsion.

In expression (1) the function $\Phi(x,y)$ is defined as

$$\Phi(x,y)=\frac{2\pi\omega(x,y)\tan\,\theta}{T}, \tag{2}$$

where $\omega(x,y)$ is the normal displacement of the investigated surface due to the deformation; θ is the angle between the projected planes of the shadows creating the system of fringes on the object and the normal to the flat surface of the object.

Figure 1 presents the optical scheme of a device for color visualization of images of deformed zones of the diffusely scattering surface during optical processing of photographs of the projected fringes in white light. A photograph of the projected fringes *3* is illuminated by a collimated beam of light emanating from a point source *1.* In this case, spectral images of the point source will be constructed in the back focal plane of the objective *4* as a result of the dispersive properties of the periodic structure of the photograph of projected fringes in all diffraction orders (except the zeroth). The images of the point source will be stretched in the plane of the visualizing slit diaphragm *5* along the dispersion line parallel to the y axis. We assume that the visualizing diaphragm *5* is oriented with the edges of the slit parallel to the x axis. In this case only the middle part of the spectral image (labeled *2* in Fig. 2), with wavelength λ_0, will be extracted by the slit from the spectral image of the point source formed by the refracted wave on those parts of the photograph *3* which correspond to the image of undeformed zones of the investigated surface. In this case, in the visualization plane (labeled *7* in Fig. 1), which is optically conjugate to the photograph *3*, the image of the undeformed zones of the investigated surface will be colored green (λ_0). If the condition

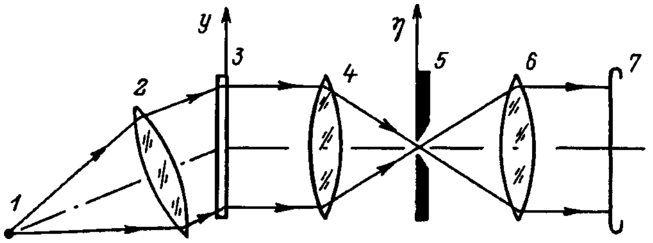

FIG. 1. Optical diagram of the device for color visualization of images of deformed surface zones: *1* — white-light point source; *2* — collimator; *3* — image of the projected fringes; *4, 6* — objectives; *5* — visualizing slit diaphragm; *7*— color visualization plane.

$$\cos\alpha_0 = n\lambda_0/T, \tag{3}$$

is fulfilled for the wave illuminating the photograph *3*, where $n=1,2,3,\dots$ and α_0 is the angle between the direction of the illuminating wave and the y axis, then, with relation (1) taken into account, it can be shown that the angles of deflection of the diffracted rays from the normal to the photograph are equal to

$$\alpha_y = \frac{\lambda_0 n}{2\pi}\frac{\partial\Phi}{\partial y}. \tag{4}$$

In expression (4) n is the diffraction order of the wave propagating along the optical axis (Fig. 1) of the receiving part of the device. It is clear from expression (4) that the spectral image *3* (Fig. 2) of the point source formed by the wave diffracted from the parts of the photograph which correspond to the image of the deformed surface zones is shifted. The linear displacement along the dispersion line is given by $\eta_\lambda \approx \alpha_y f$, where f is the focal length of the objective *4* (Fig. 1). In this case, the visualizing slit *5* will overlap the spectral segment of the image (*3* in Fig. 2) of the source with a wavelength λ different from λ_0. This causes the image of the deformed surface zone to be colored in hues other than green (λ_0). The hue of the image of the deformed zone is determined by the magnitude and sign of η_λ.

Note that in Fig. 2, which illustrates the positions of the spectral images of the light source relative to the slit of the visualizing diaphragm *1*, the images *2* and *3* are actually superimposed on each other and shifted along the η axis. With allowance for the dispersion of the diffraction grating,[17] the magnitude of η_λ can be represented in a similar way as in holographic color shadow methods,[18]

$$\eta_\lambda \approx \frac{(\lambda_0-\lambda)nf}{T}. \tag{5}$$

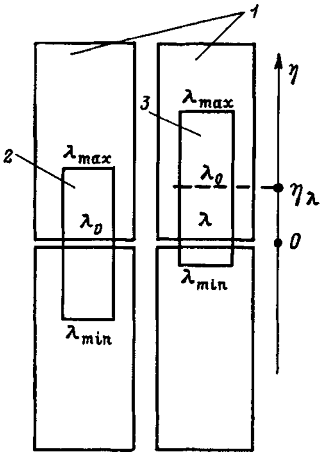

FIG. 2. Diagram of the position on the visualizing slit diaphragm of the spectral images of the point source.

This formula is valid for the propagation of a diffracted wave in the nth diffraction order only near the normal to the photograph of the projected fringes (*3* in Fig. 1). Replacing η on the left-hand side of Eq. (5) by the magnitude of the shift of the image of the light source due to deformation of the surface for the mean wavelength $[(\lambda_0+\lambda)/2]$ and taking Eqs. (2) and (4) into account, we finally obtain a working formula giving a quantitative relation between the magnitude of the deformation $\partial\omega/\partial y$ and the hue λ of the visualized zone of the investigated surface,

$$\frac{\partial\omega}{\partial y} \approx \frac{2(\lambda_0-\lambda)}{(\lambda_0+\lambda)\tan\theta}. \tag{6}$$

For a visual estimate of $\partial\omega/\partial y$ it should be borne in mind that the recordable range of hues from $\lambda_{\min}$ to $\lambda_{\max}$ is determined by the spectral sensitivity of the eye. This narrows the range of deformations $\partial\omega/\partial y$ that are measurable by the visual method. It follows from Eq. (6) that the sensitivity of the measurements and the value of $|\partial\omega/\partial y|_{\max}$ are determined by the geometry of the photographs of the projected fringes $(\tan\theta)$ and are independent of the diffraction order. It follows that it is advisable in the measurements to use the first-order diffracted wave, this being the brightest, and to choose the angle θ at which the photographs are taken with the required sensitivity and range of the measurements of $\partial\omega/\partial y$ in mind.

Note that the approach suggested here to estimating the sensitivity of the measurements and the range of the measured quantity $\partial\omega/\partial y$ and also the spectral purity of the hue in the visualized picture is similar to the approach in color methods of studying phase objects reconstructed from holograms.[18]

The use of a single photograph of the projected fringes allows one to obtain a color coding of the derivative of the displacement in only one direction — along the dispersion axis of the periodic structure of the photograph. For color coding of $\partial\omega/\partial y$ it is necessary that the projected fringes be oriented along the y axis. A reorientation of the fringes of the type (1) could be achieved by retaking the photograph using two coherent beams with a change in orientation of the projected fringes in the necessary direction, in accordance with the method of rewriting of holograms.[19] For the photographs of projected fringes such a reorientation of the fringes was used earlier to compare the macrorelief of surfaces.[20] Thus, it is possible to take a new photograph of the projected fringes with a period equal to the period of the fringes of the original photograph (1) but with the fringes oriented along the y axis. The amplitude transmittance of such a photograph is given by

$$\tau'(x,y) \sim \left\{1+\cos\left[\frac{2\pi x}{T}+\Phi(x,y)\right]\right\}^{-\gamma/2}. \tag{7}$$

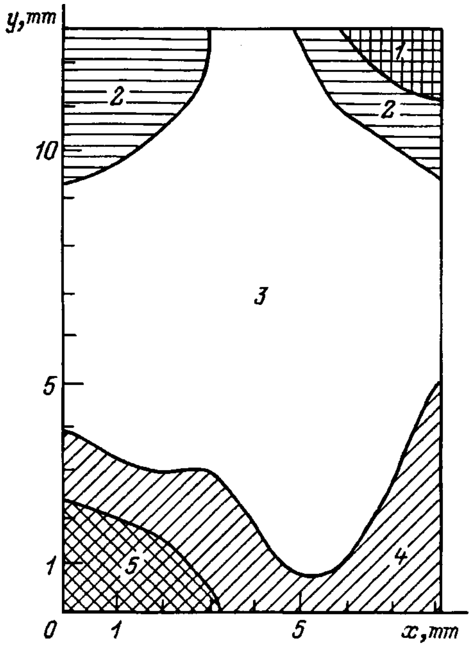

FIG. 3. Color visualization of deformed zones of an aluminum plate. Zones are colored in various hues: *1* — red, *2* — orange, *3* — yellowish-red, *4* — blue, *5* — violet.

In the optical processing of a photograph of this type (7) in the scheme depicted in Fig. 1, deformed image zones characterized by the derivative $\partial\omega/\partial x$ are visualized. A working formula analogous to (6) is used to obtain a quantitative estimate of $\partial\omega/\partial x$ from the hue.

The method of color coding of deformed zones was experimentally verified by visualizing deformations of a flat aluminum plate with dimensions 13×8 mm, which was subjected to strong plastic deformations. Figure 3 depicts the plate observed in different colors in the plane *7* (Fig. 1) of the color visualization device. The hue in the observed image characterizes the degree of deformation $\partial\omega/\partial y$ of the surface of the plate. The color image was adjusted so that the images of the undeformed zones were colored yellow-green. In Fig. 3 the area of the plate is divided into five hues. From the mean wavelength of each of the five hues in the visualized image, the value of $\partial\omega/\partial y$ was determined using the working formula (6). The mean wavelength of each hue was estimated visually by comparing with standard hues as has been done in color shadow methods.[18] The relative error of determination of the mean wavelength in the visualized image was determined both by the spectral purity of the hue and by subjective idiosyncrasies of the eye of the experimenter. In a specific case, the relative error of the estimated mean wavelength did not exceed 25%; however, it could be substantially reduced, for example, by the use of a receiver with better spectral resolution.

In conclusion, I would like to remark that the method of visualizing deformations of diffusely scattering surfaces described above is somewhat less accurate than the method developed earlier.[10] However, in contrast to Ref. 10, the present method allows one simultaneously to visualize all of the deformed zones of the investigated surface.

This work was performed with the support of the Ministry of National Education of the Republic of Belarus.

[1] C. M. Vest, *Holographic Interferometry* (Wiley, New York, 1979).

[2] A. G. Kozachok, *Holographic Methods of Study in Experimental Mechanics* (Mashinostroenie, Moscow, 1984), 175 pp.

[3] K. Patorski and M. Kujawińska, Appl. Opt. **24**, 3041 (1985).

[4] K. Patorski, D. Post, R. Cramek, and Guo Yifan, Appl. Opt. **26**, 1977 (1987).

[5] R. K. Erf (Ed.), *Holographic Nondestructive Testing* (Academic Press, New York, 1974).

[6] S. N. Koreshev and A. G. Seregin, Opt. Spektrosk. **77**, 991 (1994) [Opt. Spectrosc. **77**, 893 (1994)].

[7] A. M. Lyalikov, Opt. Spectrosc. **74**, 865 (1993) [Opt. Spectrosc. **74**, 516 (1993)].

[8] A. M. Lyalikov, Opt. Spectrosc. **77**, 959 (1994) [Opt. Spectrosc. **77**, 863 (1994)].

[9] A. M. Lyalikov, Zh. Tekh. Fiz. **67**(3), 67 (1997) [Tech. Phys. **42**, 313 (1997)].

[10] A. M. Lyalikov, Zh. Tekh. Fiz. **68**(3), 74 (1998) [Tech. Phys. **43**, 332 (1998)].

[11] J. E. O'Hare and J. D. Trolinger, Appl. Opt. **8**, 2047 (1969).

[12] A. F. Belozerov and N. M. Spornik, Opt. Mekh. Promst., No. 3, 9 (1971).

[13] I. S. Zeĭlikovich and N. M. Spornik, Opt. Spektrosk. **46**, 393 (1979) [Opt. Spectrosc. **46**, 217 (1979)].

[14] I. S. Zeĭlikovich, A. M. Lyalikov, and N. M. Spornik, Opt. Spektrosk. **62**, 659 (1987) [Opt. Spectrosc. **62**, 392 (1987)].

[15] H. Bartelt, S. K. Case, and R. Hauck, in *Applications of Optical Fourier Transforms*, edited by H. Stark (Academic Press, New York, 1982; Radio i Svyaz', Moscow, 1988, pp. 472–508).

[16] S. Jutamulia, T. W. Lin, and F. T. S. Yu, Opt. Commun. **58**, 78 (1986).

[17] M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon Press, Oxford, 1969; Mir, Moscow, 1973, 719 pp.).

[18] A. K. Beketova, A. F. Belozerov, A. N. Berezkin *et al.*, *Holographic Interferometry of Phase Objects* [in Russian] (Leningrad, 1979), 232 pp.

[19] I. S. Zeĭlikovich and S. A. Pul'kin, Opt. Spektrosk. **53**, 588 (1982) [Opt. Spectrosc. **53**, 349 (1982)].

[20] A. M. Lyalikov, Opt. Spektrok. **80**, 849 (1996) [Opt. Spectrosc. **80**, 764 (1996)].

Translated by Paul F. Schippnick

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Development of a system of anticoronal shields for the ion beam injector and accelerator of a diagnostic complex for plasma physics research

M. M. Rezinkina, O. S. Nedzel'skiĭ, S. M. Khrebtov, and O. L. Rezinkin

(Submitted October 1, 1997)
Zh. Tekh. Fiz. **68**, 106–109 (November 1998)

A system of anticoronal shields is proposed, designed, and tested. The system is found to work efficiently. © *1998 American Institute of Physics.* [S1063-7842(98)02111-4]

Diagnostics of the parameters of a hot plasma using a heavy-ion probe beam is presently one of the more forward-looking methods of study of such objects. Heavy-ion beam injectors and electrostatic accelerators are used for this purpose. In order to ensure efficient functioning of the given apparatus, it is necessary to prevent the appearance of a corona discharge from the metallic elements of the injector and accelerator tube, which have a positive voltage on them of $0-200$ kV during use. The sharp edges and small radii of curvature of these elements give rise to corona discharges from their edges. The appearance of a corona discharge leads to instability and a nonequilibrium distribution of the accelerating potentials, overloading of the voltage source, the appearance of intense electromagnetic interference, ionization and ozonation of air in and around the apparatus, and a lowering of the reliability of operation of all the systems and apparatus of the analyzer.

A condition for the appearance of a corona near the surface of metallic elements under a voltage (below we will refer to them simply as electrodes) is the presence of an electric field exceeding the initial electric field E of the corona. Under standard atmospheric conditions E is of the order of 30 kV/cm.[1,2] Preliminary estimates and also experience with such designs show that a corona discharge will take place near their surfaces. Among the most effective means of dealing with this phenomenon is a setup of shielding electrodes.[3] On the basis of known engineering solutions[2–4] and experience with the development of high-voltage equipment, we chose a design of the system of field-forming shields (SFS) whose main elements are depicted in Fig. 1. The shield of the ion injector *1*, made in the form of a hood with radius of curvature 80 mm, and the flange shield *2* completely shield all the electrodes of the injector unit. The first and second intermediate shields *3* and *4*, which are at potentials of 140 and 80 kV, respectively, lower the maximum electric field levels on the gradient rings *5*. The gradient rings *5* lower the electric field on the resistors *6*, which are located between the rings along the length of the accelerating tube *7* and serve to distribute the accelerating voltage.

The ion injector shield *1* is found at a maximum potential of $+200$ kV during operation. Regarding its design, the shield has three main parts: the injector hood *1*, flange *8*, and flange shield *2* (see Fig. 1). The hood *1* has a cylindrical part and a face part with rounded edges. The center of the face part has an opening for the high-voltage cable *9*. The nominal mass of the shield *1* is 1.4 kg. The first intermediate shield *3* of the SFS is designed to lower the electric field intensity near the surface of the gradient rings, which are found at potentials of $+190-140$ kV during use. The shield is at a potential of $+140$ kV. It consists of a toroidal, a cylindrical, and a conical part. During assembly the shield is mounted to one of the gradient rings *5*. The nominal mass of the first intermediate shield is 5.7 kg.

The second intermediate shield *4* is designed to lower the electric field intensity near the surface of the gradient rings, which are at potentials of $+130-80$ kV during use. The shield is at a potential of $+80$ kV. This shield has a toroidal and a conical part. During assembly, this shield is mounted to one of the gradient rings *5*. The nominal mass of the second intermediate shield is 4.48 kg. Each gradient ring *5* consists of a torus fabricated from aluminum thin-walled tubing of diameter 10 mm and welded to it along the inner surface of the aluminum ring of thickness 1 mm. The ring has elements mounting it to the electrodes of the accelerating tube *7* (Fig. 1). Mounted to the surface of the ring by rivets on both sides are fringes to which KÉV-1 resistors are soldered (position *6* in Fig. 1).

To estimate the efficacy of the SFS design described above, we calculated the electric field distribution in the regions of maximum field. Preliminary estimates enabled us to select two such dangerous regions: region *A* and region *B* (Fig. 1). Since the investigated zones possess axial symmetry, we used cylindrical coordinates. This allowed us to solve the problem of calculating the field in two-dimensional form. We assumed that **E** depends only on the radial r and the

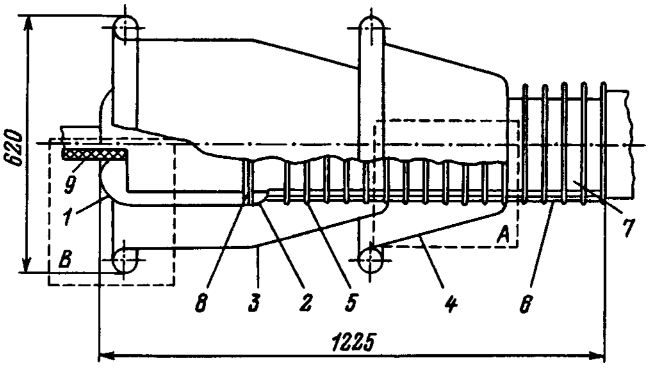

FIG. 1. Diagram of the system of anticoronal shields.

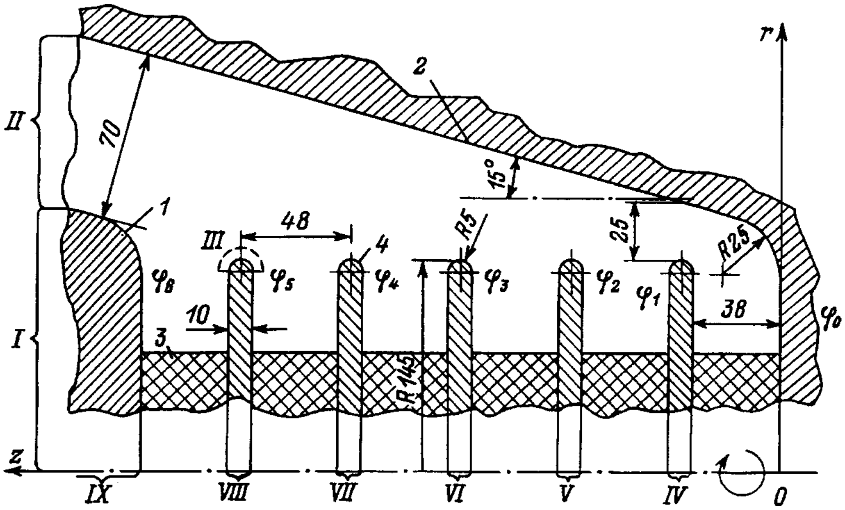

FIG. 2. Axial cross section of the first design region: *1* — shield *1*; *2* — shield *2*; *3* — insulator ring; *4* — gradient ring.

azimuthal z coordinates. In view of the complicated geometrical shapes of the electrodes and the SFS, we utilized the method of finite differences.

Let us consider an axial cross section of the zone in question. A nonuniform rectangular mesh was imposed on the computational region, with mesh lines parallel to the z axis: $r_1=0$; $r_2=\Delta r_1$; $r_3=r_2+\Delta r_2,\ldots$; $r_i=r_{i-1}+\Delta r_{i-1}$, $\ldots$; $r_{NR}=r_{NR-1}+\Delta r_{NR-1}$, and mesh lines parallel to the r axis: $z_1=0$; $z_2=\Delta z_1$; $z_3=z_2+\Delta z_2,\ldots$; $z_j=z_{j-1}+\Delta z_{j-1},\ldots$; $z_{NZ}=z_{NZ-1}+\Delta z_{NZ-1}$ (where NR is the number of divisions along the r axis, and NZ is the number of divisions along the z axis; Δr_i and Δz_j are the step sizes of the mesh in r and z, respectively). The quantities Δr_i and Δz_j were determined from the required accuracy of the calculation so as to take account of field distortions in all the elements of the system. Toward this end, the step size in the inhomogeneity region, for example, of the gradient ring *5*, was chosen to be 8–10 times smaller than the length of the given inhomogeneity. The boundary conditions were determined by the type of system. Noting that the voltage on the electrodes is constant, we write down the following equation for each mesh point of the computational mesh:

$$\int_S \gamma\cdot E_n ds=0,$$

where S is a surface encompassing the mesh point in such a way that it divides the distances between neighboring mesh points in half; the subscript n denotes the projection of the electric field vector $\mathbf{E}$ on the surface normal.

We now express $\mathbf{E}$ in the latter expression in terms of the values of the potential $\varphi_{i,j}(r,z)$ at the mesh points of the computational mesh. Finally, we write it in difference form for the (i,j)-th cell

$$\Lambda_r\varphi_{i,j}+\Lambda_z\varphi_{i,j}=0,$$

where

$$\Lambda_r\varphi_{i,j}=\varphi_{i-1,j}\cdot AR_{i,j}-\varphi_{i,j}(AR_{i,j}+BR_{i,j})+\varphi_{i+1,j}\cdot BR_{i,j},$$

$$\Lambda_z\varphi_{i,j}=\varphi_{i,j-1}\cdot AZ_{i,j}-\varphi_{i,j}(AZ_{i,j}+BZ_{i,j})+\varphi_{i,j+1}\cdot BZ_{i,j},$$

$$AR_{i,j}=\frac{1}{\Delta r_{i-1}}\left(r_i-\frac{\Delta r_{i-1}}{2}\right)\left(\frac{\Delta z_{j-1}}{2}\gamma_{i-1,j-1}+\frac{\Delta z_j}{2}\gamma_{i-1,j}\right),$$

$$BR_{i,j}=\frac{1}{\Delta r_i}\left(r_i+\frac{\Delta r_i}{2}\right)\left(\frac{\Delta z_{j-1}}{2}\gamma_{i,j-1}+\frac{\Delta z_j}{2}\gamma_{i,j}\right),$$

$$AZ_{i,j}=\frac{1}{\Delta z_{j-1}}\left[\left(r_i+\frac{\Delta r_i}{4}\right)\frac{\Delta r_i}{2}\gamma_{i,j-1}+\left(r_i-\frac{\Delta r_{i-1}}{4}\right)\times\frac{\Delta r_{i-1}}{2}\gamma_{i-1,j-1}\right], \tag{1}$$

$$BZ_{i,j}=\frac{1}{\Delta z_j}\left[\left(r_i+\frac{\Delta r_i}{4}\right)\frac{\Delta r_i}{2}\gamma_{i,j}+\left(r_i-\frac{\Delta r_{i-1}}{4}\right)\frac{\Delta r_{i-1}}{2}\gamma_{i-1,j}\right], \tag{2}$$

$\gamma_{i,j}$ is the conductivity of the cell, whose vertices are the mesh points (i,j), $(i+1,j)$, $(i,j+1)$, and $(i+1,j+1)$.

The above equation was solved by an alternaing-directions iterative method using a program written in FORTRAN-77 to be run on an IBM PC. The calculations were similar to those of Ref. 5.

Figure 2 displays an axial cross section of the first computational region (Fig. 1,A). In the calculation we imposed the following boundary conditions with respect to r. The conditions at $r=0$ $(i=1)$ were: $\varphi_{1,1}=\varphi_0=140$ kV at $z=0$; $\varphi_{1,j}=\varphi_1=150$ kV for z belonging to region *IV*; $\varphi_{1,j}=\varphi_2=160$ kV for z belonging to region *V*; $\varphi_{1,j}=\varphi_3=170$ kV for z belonging to region *VI*; $\varphi_{1,j}=\varphi_4=180$ kV for z belonging to region *VII*; $\varphi_{1,j}=\varphi_5=190$ kV for z belonging to region *VIII*; $\varphi_{1,j}=\varphi_6=200$ kV and for z belonging to region *IX*. For $z>0$ and outside regions *IV–IX*, as a consequence of the axial symmetry of the system, we used homogeneous boundary conditions of the second kind (Neumann conditions): $\partial\varphi/\partial r=0$. The condition at $r=r_{\max}$ was $\varphi_{NR,j}=140$ kV. The boundary conditions with respect to z were as follows: $-\varphi_{i,1}=\varphi_0=140$ kV at $z=0$ $(j=1)$, $\varphi_{i,NZ}=140$ kV at $z=z_{\max}$ $(j=NZ)$ for r belonging to region *I*; and $\partial\varphi/\partial z=0$ for r belonging to region *II*. The potential of the point $(NR,\ NZ)$ was $\varphi_{NR,NZ}=140$ kV.

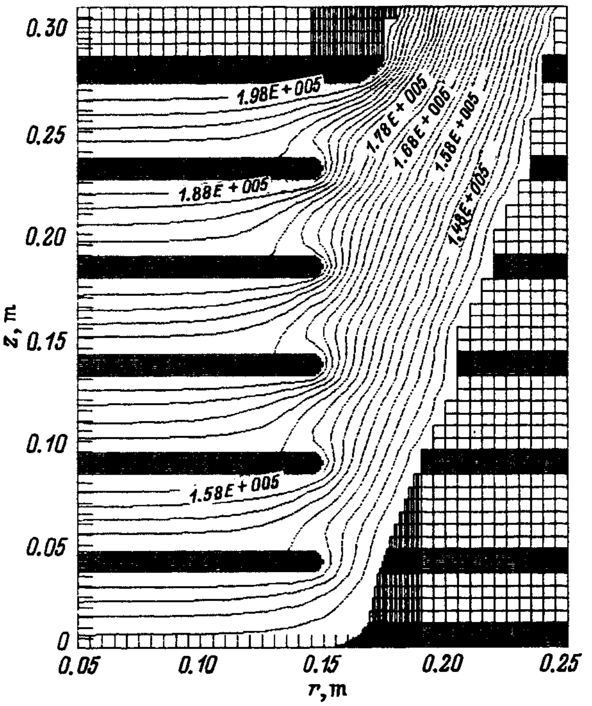

FIG. 3. Distribution of the potential in the first calculation region.

Calculated distributions of the electric field in region *A* of Fig. 1 are shown in Fig. 3 and summarized in Table I.

Figure 3 plots equipotentials in an axial cross section of the system. The numerical values labeling the equipotential lines are given in volts. Table I lists the maximum fields reached in the zones that are the most dangerous from the point of view of the appearance of a corona. As can be seen from Table I, the maximum field level does not exceed 16 kV/cm, which speaks of reliable protection from corona in that region.

Figure 4 shows an axial cross section of the second calculational region (region *B* in Fig. 1). In the calculations we imposed the following boundary conditions with respect to r: $\varphi_{1,j}=\varphi_2=200$ kV for $r=0$ on the symmetry axis of the system ($i=1$) for z belonging to region *I*, and $\partial\varphi/\partial r=0$ for z belonging to region *IV*. The potential at the point $(1, NZ)$ was $\varphi_{1,NZ}=0$. The boundary conditions for $r=R_{\max}$ ($i=NR$) were $\varphi_{NR,j}=0$. The boundary conditions with respect to z were as follows: $\varphi_{i,1}=\varphi_2=200$ kV for r belonging to region *I*, and $\varphi_{i,1}=\varphi_1=140$ kV for r belonging to region *II*. Outside these zones we used the homogeneous boundary condition of the second kind (Neumann) $\partial\varphi/\partial z=0$.

The potential distribution in such a system is shown in Fig. 5, and the maximum field values are listed in Table I. To start with, the design of the given electrode system was somewhat different. Calculations carried out for an electrode system similar to that shown in Fig. 4, but with smaller radii of curvature (the radius of curvature of region *VI* was 70 mm, and region *VII* was not curved), showed that the maximum field levels in such a system reach 25 kV/cm. Fields of such intensities under certain conditions can give rise to a corona. The geometry of the system was modified to lower the field levels (Fig. 4). As can be seen from the table, the use of this SFS geometry lowered the maximum field levels to

TABLE I. Maximum electric field levels in zones of the system of anticoronal shields.

Zone	Fig. 2		Fig. 4	
	Shield *I*	Zone *III*	Zone *VI*	Zone *V*
$\lvert E_{\max}\rvert$ [kV/cm]	16	15.5	14.6	14.4

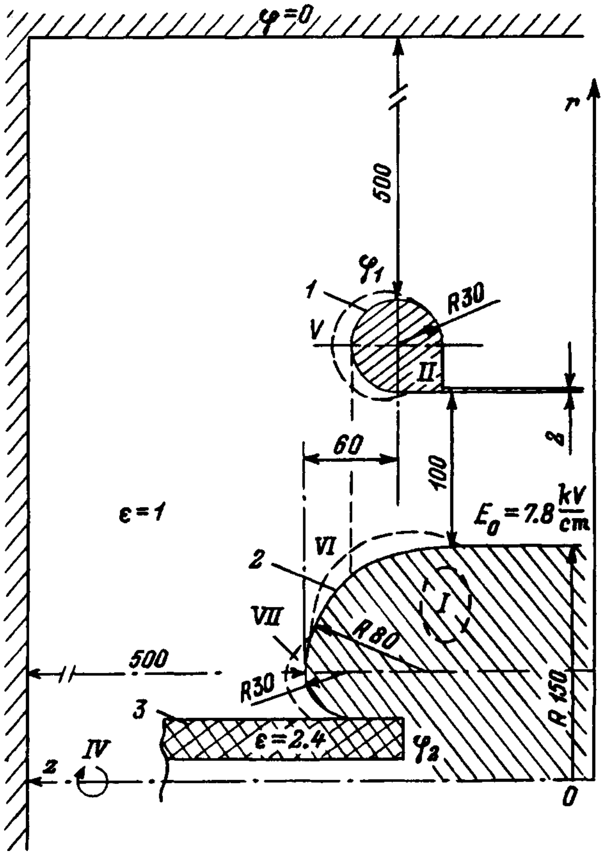

FIG. 4. Axial cross section of the second calculation region: *1* — shield, *2* — hood, *3* — cable.

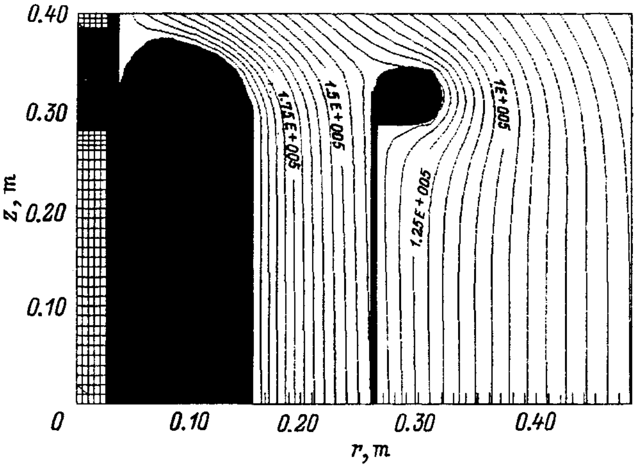

FIG. 5. Distribution of the potential in the second calculation region.

14.6 kV/cm, which makes it possible to avoid the appearance of a corona on the electrode and shield surfaces.

To summarize, we have developed an SFS design that reliably suppresses corona from elements of the system under voltage. This makes it possible to increase the reliability and stability of operation of the accelerator, increase the duty factor of the voltage source, and reduce the electromagnetic interference and ozone concentration in the work area.

This work was carried out within the purview of a contract between the Institute of Plasma Physics (Khar'kov) and the Center for Research on Energy, the Environment, and Technology, Madrid [Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas (CIEMAT) (Madrid)] for the Period 1996–2000.

[1] G. N. Aleksandrov, V. L. Ivanov, and V. E. Kizevetter, *Dielectric Strength of External High-Voltage Insulation* [in Russian] (Énergiya, Moscow, 1969), 238 pp.
[2] M. V. Kostenko (Ed.), *High-Voltage Technology* [in Russian] (Vysshaya Shkola, Moscow, 1973), 527 pp.
[3] V. M. Tubaev, Candidate's Dissertation [in Russian], Kharkov (1965).
[4] G. S. Kuchinskiĭ, V. E. Kizevetter, and Yu. S. Pintal', *Insulation of High-Voltage Installations* [in Russian] (Énergoatomizdat, Moscow, 1987), 368 pp.
[5] M. M. Rezinkina and O. L. Rezinkin, Élektrichestvo, No. 7, 62 (1995).

Translated by Paul F. Schippnick

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

BRIEF COMMUNICATIONS

Plasma focus as a current switch for a capillary discharge

É. Yu. Khautiev

Sukhumi Physicotechnical Institute, Sukhumi, Abkhazia

P. S. Antsiferov, L. A. Dorokhin, K. N. Koshelev, and Yu. V. Sidel'nikov

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow District, Russia

(Submitted December 30, 1996)
Zh. Tekh. Fiz. **68**, 110–113 (November 1998)

Experiments are described in which a plasma focus is used simultaneously as an inductive store and a current switch. The obtained rates of current growth on a load of 0.01 Ω is 10^{12} A/s, and the maximum values of the switched current lie in the range 50−100 kA. The technique is seen as promising for employing a capillary discharge as a source of laser medium for the soft x-ray region. © *1998 American Institute of Physics.* [S1063-7842(98)02211-9]

INTRODUCTION

The increased interest in the use of a capillary discharge as a possible source of a medium with population inversion on transitions in the soft x-ray range[1,2] has stimulated the search for new forms of its practical implementation. Although the total energy of such a discharge can be very small (10−100 J), its realization is by no means a trivial task. The main difficulty here has to do with the necessity of providing a sufficiently steep current rise (10−100 kA over times of the order of 10−100 ns) in order to eliminate the influence on the discharge of the evaporating walls of the capillary.

As is well known, despite their much greater energy capacity in comparison with capacitive energy stores, inductive stores have not met with as wide use, since switching them over to a load requires interruption of the current circuit.[3,4] A number of devices developed toward this end, such as plasma and explosive opening switches are in themselves quite complicated devices and do not always provide the required current rise at the load. In this paper we report the use of a very widely used device, namely a plasma focus, in the simultaneous roles of an inductive energy store and a device for switching the current to the load.

The main idea here is the following: after the current layer detaches from the insulator and reaches the discharge axis, the anode–cathode electrode system of the plasma focus is an inductance with stored energy W. Further development of the discharge in certain regimes leads to a sizable jump in the active resistance in the pinched plasma column. This jump gives rise to an overvoltage, which can throw part of the energy W over into the load circuit, connected through a spark gap in parallel with the anode–cathode circuit of the plasma focus, over times much shorter than the discharge time in the main discharge circuit.

EXPERIMENT

An overall diagram of the experimental device is shown in Fig. 1. It contains two main parts: the plasma focus proper and the load unit. A plasma focus of the maser type was used, having an outer electrode (cathode) diameter of 70 mm, an inner electrode (anode) diameter of 26 mm, and a total length of the coaxial electrode system of 110 mm. The porcelain insulator had a length of 30 mm and a diameter of 26 mm. The main discharge capacitance C was 10 μF, and the working voltage was 25 kV. The total inductance of the discharge circuit up to the vacuum chamber (inductance of the main capacitor, the system of power cables, and the air-filled spark gap G_1) is estimated as $L_i = 0.11$ μH. The time of the first half period was 3.2 μs. The maximum current in the main circuit (with the capillary disconnected) was about 200 kA.

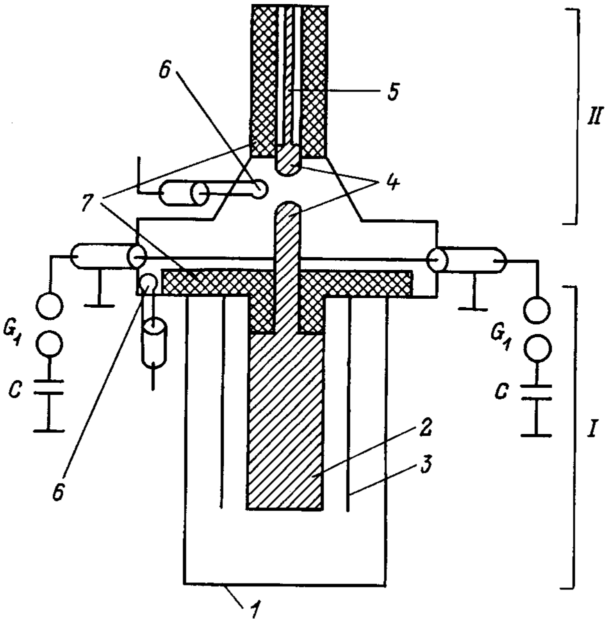

FIG. 1. Diagram of the whole experimental setup: *I* — plasma focus, *II* — load unit; *1* — vacuum chamber, *2* — anode, *3* — cathode, *4* — spark gap G_2, *5* — load, *6* — magnetic probe, *7* — insulator.

The process of current switching was investigated with a load consisting of a thin-walled stainless-steel tube with outer diameter 4 mm, working length 100 mm, and resistance $R_1 = 0.01\ \Omega$. Regarding the overall design of the circuit, the stainless-steel tube is coaxial with the electrode system of the plasma focus and is separated from the anode by an air discharge gap of 10−13 mm (the discharge gap G_2 in Fig. 1). The return current lead of the resistance consists of a cylinder with an inner diameter of 6 mm and is isolated over its entire length from the load tube. The inductance of the load unit together with the discharge gap G_2 is estimated as $L_2 = 0.03\ \mu$H. Two inductive pickups allowed us to record the current in the plasma focus circuit and the load circuit.

The numerical model that was used to analyze the behavior of the system was based on the equivalent electrical circuit depicted in Fig. 2. The model assumed that the total inductance of the discharge chamber L_3 varies smoothly, as a consequence of motion of the current envelope, from 0.01 μH at the onset of the discharge to 0.03 μH at the end of the first quarter period. The key element of the model is the active resistance of the discharge R_2. Raising this resistance at a certain stage of pinching of the plasma leads to a well-known phenomenon—the appearance of an overvoltage in the discharge. The magnitude of this overvoltage U for setups similar in their parameters to ours is on the order of 100 kV and higher. For a current of about 200 kA this corresponds to a maximum value of $R_2 \approx 0.5-1\ \Omega$.

The spark gap G_2 breaks down when the overvoltage reaches a certain value. From that time onward, a current develops in the load circuit that is directed opposite the main current. The rate of growth of the current in the loaded (capillary) circuit is equal to its rate of decrease in the main circuit (L_2, R_2) of the plasma focus. The derivative of this current can be estimated from the relation $dI/dt \approx U/(L_2 + L_3)$ and has a value $\geqslant 10^{12}$ A/s for the parameter values discussed. The total value of the switched current also depends on the existence time of the high resistance R_2. For characteristic times of the order of several tens of nanoseconds, one can hope for a total current in the load (capillary) somewhere in the range from 50 to 100 kA. An exact solution of the electrical-engineering equations for the equivalent circuit in Fig. 2 is described below. As comparison with the experimental observations shows, the equivalent circuit provides a good description of the overall behavior of the current switch.

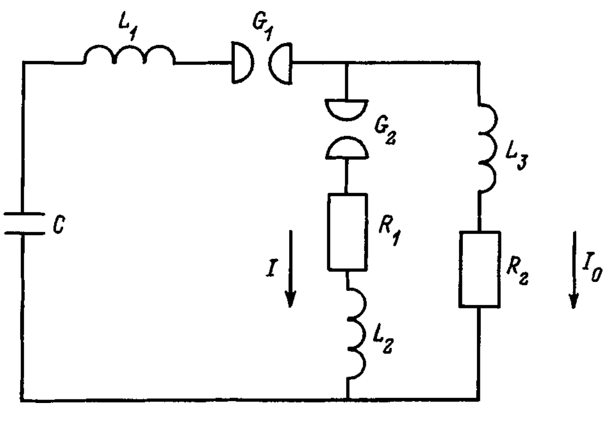

FIG. 2. Equivalent circuit of the main electrical circuit.

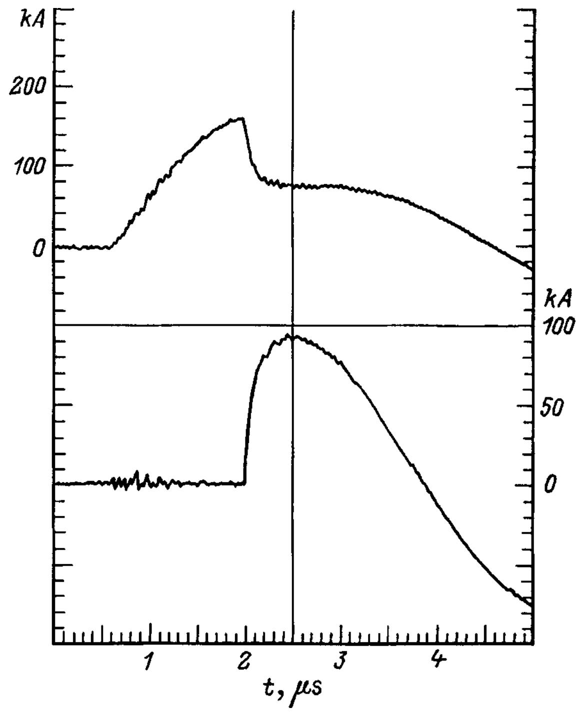

FIG. 3. Oscillograms of the current through the plasma focus (upper graph) and through the load (lower graph).

RESULTS AND DISCUSSION

In the present work we have investigated the switching of a discharge current to load using hydrogen, helium, and neon as the working gas, with pressures ensuring the arrival of the current envelope at the discharge axis at the instant of maximum discharge current. Figure 3 presents experimental oscillograms of the current flowing through the discharge and load for the case of successful switching (10−20% of the total number of discharges, see below). It can be seen that during a time of the order of 50 ns a current of 50 kA is successfully switched onto the load, i.e., a rate of current rise of 10^{12} A/s is achieved.

The experimental oscillograms were analyzed with the help of a numerical model of the equivalent circuit (Fig. 2). Results of numerical calculations of the time dependence of the corresponding currents are plotted in Fig. 4.

At the initial instant of time $t=0$ the capacitance C is charged to 25 kV and all currents in the circuit are equal to 0. At this same time breakdown occurs at the main spark gap G_1 and a current starts to develop through the plasma focus. The time t_0 at which the maximum value of the active resistance R_2 is reached was determined from the oscillograms of the current and corresponded to the onset of current switching. The time dependence of the resistance R_2 is given by the formula

$$R_2 = R_0 \exp(-(t-t_0)^2/\delta t^2). \tag{1}$$

The spark gap G_2 switched on when the overvoltage reached 30 kV. The development time of the discharge in

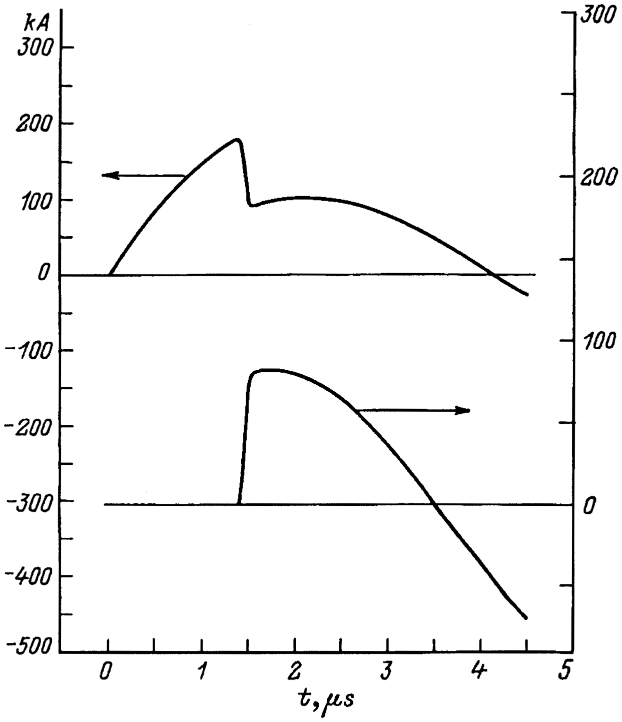

FIG. 4. Calculated curves of the time dependence of the current through the plasma focus (upper graph) and through the load (lower graph).

air-filled spark gaps ordinarily does not exceed 10 ns and is not taken into account in the present estimates. Values of the drop in the current through the discharge and the corresponding jump in current at the load depend mainly on the product $R_0 \delta t$ (provided $R_1 < R_0$), which under our conditions is approximately equal to 0.02 $\Omega \cdot \mu$s for a current jump of 50 kA. This latter fact can be understood if we observe that the current increment in the $R_1 - L_2$ circuit is determined by the equation $dI/dt = (L_2 + L_3)^{-1} \cdot I_0 \cdot R_2(t)$, and if the current through the plasma focus I_0 varies weakly during the current throw time (this assumption, generally speaking, is poorly fulfilled, and therefore the dependence in question is not absolutely exact), then we may use the estimate

$$I(t) \cong I_0 \cdot (L_2 + L_3)^{-1} \int_0^t R_2(t)\, dt. \tag{2}$$

Here the integral can always be estimated as the product of some characteristic resistance and the time of the process, i.e., as was already noted, the value of the switched current under these conditions does not depend on the details of the time history of R_2. Bearing in mind that the total switching time is a quantity of the order of δt, we may conclude that the average value of dI/dt at the load is determined by R_0. Note that the details of the current rise in the experimental oscillograms indicate a time dependence of this resistance that is more complicated than is represented in formula (1).

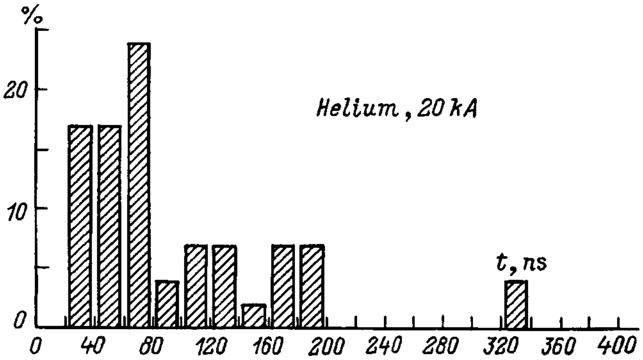

FIG. 5. Statistics of the rise time of the current in the load to values of the order of 20 kA with helium as the working gas in the plasma focus.

The discharge of a plasma focus is, generally speaking, a poorly reproducible physical object. The process of using it to switch the current onto the load is therefore characterized by some statistics. Figure 5 plots results in the form of a histogram of statistical processing of a series of experiments (50 discharges) using helium as the working gas in the plasma-focus discharge. For each experimental oscillogram we determined the time it took the current through the load to reach 20 kA. The histogram plots percent distributions of the number of discharges in which the current switched onto the load reached a given value between the times t and $t + \Delta t$ ($\Delta t = 20$ ns). The histogram gives a picture of the mean and limiting characteristics of the current rise at the load. Thus, in 17% of the total number of discharges a current growth rate of 0.5×10^{12} to 1×10^{12} A/s was obtained. From the point of view of obtaining the shortest rise times it is preferable to use He or Ne instead of H_2 as the working gas. The pressure of the working gas does not have a substantial effect on the current switching process as long as the singularity of the discharge current through the plasma focus remains in the region of its maximum (between 1 and 2 μs). Note that no effort was undertaken to optimize the plasma focus itself from the point of view of the switching process and the given statistics could therefore possibly be improved.

This work was carried out with the support of the Russian Fund for Fundamental Research (Project No. 95-02-04495a).

[1] C. Steden and H.-J. Kunze, Phys. Lett. A **151**, 534 (1990).

[2] J. J. Rocca, O. D. Cortazar, B. Shapiro *et al.*, Phys. Rev. E **47**, 1299 (1993).

[3] G. A. Mesyats, *Generation of High-Power Nanosecond Pulses* [in Russian] (Sov. Radio, Moscow, 1974).

[4] B. M. Koval'chuk, Yu. A. Kotov, and G. A. Mesyats, Zh. Tekh. Fiz. **44**, 215 (1974) [Sov. Phys. Tech. Phys. **19**, 136 (1974)].

Translated by Paul F. Schippnick

Internal friction in directed-crystallization (Cu–Sn)–Nb alloys

V. M. Arzhavitin and V. Ya. Sverdlov

Kharkov Physicotechnical Institute National Science Center, Ukraine Academy of Sciences, 310108 Kharkov, Ukraine

(Submitted February 19, 1997)
Zh. Tekh. Fiz. **68**, 114–117 (November 1998)

The distinctive features of the low-frequency internal friction $Q^{-1}(T)$ of (Cu–Sn)–Nb composites at high temperatures (up to 400 °C) are investigated for strains in the range $10^{-5}-10^{-4}$. Considerable hysteresis of $Q^{-1}(T)$ in the heating–cooling cycle is recorded, including the presence of a minimum at ~175 °C when the sample is heated to 400 °C and two peaks P_2 (at 280 °C) and P_1 (at ~100 °C) when the sample is cooled from 400 °C. The activation energy of the anomalous internal friction background (up to 175 °C), the oxygen diffusion parameters, and the oxygen concentration in the niobium fibers (all of which govern the peak P_2) are calculated, and the value and temperature dependence of the yield point of the bronze matrix (which govern the peak P_1) are estimated. © *1998 American Institute of Physics.* [S1063-7842(98)02311-3]

INTRODUCTION

In a number of composites obtained by directed crystallization of alloys it is possible to avoid degradation of the strength properties right up to pre-melting temperatures. The reason for this encouraging result lies in the relative perfection (semicoherence) of the surfaces that form between the phases. The high heat resistance of such composites makes them useful for making the nozzle vanes of gas turbines, combustion chambers, heat shields, and other thermally loaded components. However, when choosing a composite for operation in nonstationary thermal regimes it is necessary to take into account the deleterious effect of interphase thermal stresses on the directed microstructure. These stresses can be large enough to cause extremely undesirable disruption in the composites, e.g., fragmentation of the fibers and microplastic deformation of the matrix.[1]

A vast amount of information about elementary microplasticity events in metals and alloys has been obtained by studies of low-frequency internal friction, since measurements of this phenomenon are essentially direct experiments in microdeformation.[2] However, at this time there is a noticeable lack of studies based on amplitude-dependent internal friction that address the problem of microplasticity of metal composites. It is this circumstance that dictates the goal and nature of this work, i.e., the analysis of temperature spectra of low-frequency internal friction obtained under thermal-cycling conditions in directed-crystallization (Cu–Sn)–Nb composites.

SAMPLES AND MEASUREMENT METHOD

Ingots of the ternary alloys (Cu–13%Sn)–30%Nb and (Cu–8%Sn)–30%Nb were obtained by directed crystallization in a high temperature gradient (~40 K/mm). The rate of cooling during hardening was ~100 K/s. This high rate of cooling is necessary in order to prevent the formation of the brittle compound Nb_3Sn during directed crystallization, and to ensure plastic deformation of the ingots. Initially, these Cu–Nb–Sn ingots had a typical composite microstructure in the form of a bronze matrix with distributed niobium dendrites situated with their first-order axis extending along the longitudinal axis of the ingot. The transverse size of these niobium dendrites was 2–3 μm (Fig. 1). The ingots were deformed into wires by drawing with intermediate anneals at 350 °C/h in a vacuum after every 30–40% deformation. In the course of this deformation the niobium dendrites were stretched into long ribbon-shaped fibers.

Measurements of $Q^{-1}(T)$ for these samples were made by observing the free decay of torsional and resonant bending oscillations with strain amplitudes in the range $10^{-4}-10^{-5}$ at frequencies 1–70 Hz in a vacuum of $\sim 10^{-1}$ Pa. The range of thermal cycling was limited to temperatures of 20–400 °C in order to avoid the formation of niobium stannide (Nb_3Sn), the homogeneity region of which in the Nb–Sn system is still in need of some refinement.[3]

RESULTS AND DISCUSSION

In our torsion experiments on (Cu–13%Sn) wires of diameter 0.5 mm we recorded considerable hysteresis in $Q^{-1}(T)$ over the heating–cooling cycle. In particular, the internal friction exhibited a minimum (at 200 °C) during the heating segment of the cycle and two peaks P_2 (at 280 °C) and P_1 (at 100 °C) as the sample was cooled down from 400 °C (see Fig. 2, curves *1* and *2*). We also found that the temperature range (20 − 175 °C) within which the internal friction decreases overlaps the region where the so-called ''plasticity dip'' of tin-containing bronzes occurs.[4] In this range, the decrease in the internal friction can be approximated by an inverse Arrhenius relation $Q^{-1}=A\exp(U/RT)$, where R is the gas constant, A is a constant, and the energy parameter $U=0.1$ eV is comparable to the activation energy for the hysteretic motion of dislocations in metallic materials.[5]

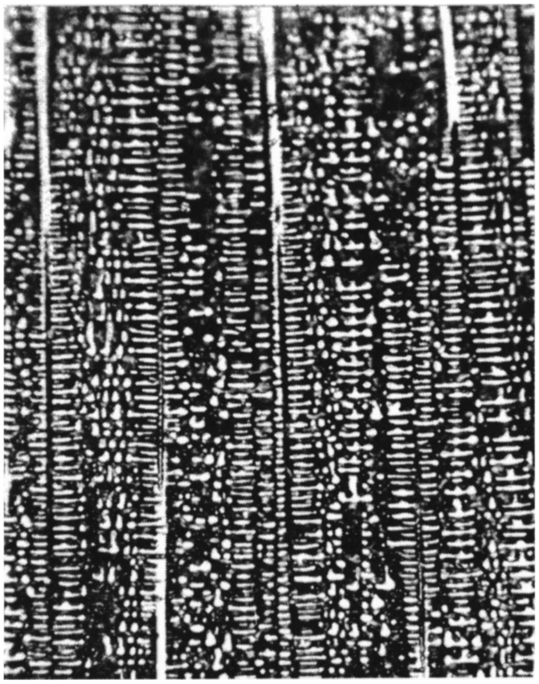

FIG. 1. Microstructure of the directed-crystallization alloy (Cu–13%Sn)–30%Nb (longitudinal section).

The peak P_2, which is not observed in pure copper, appears when 0.5-mm diameter wires made of the Cu–25%Nb composite are subjected to torsion (Fig. 2, curve *3*), and is already discernible during the heating portion of the cycle. This leads us to associate this peak with the Nb fibers. The thermal-activation parameters of the compound peak P_2 (the inset to Fig. 2) were calculated from resonance measurements for a (Cu–8%Sn)–30%Nb sample according to the temperature position of the corresponding internal friction peak. The values obtained, viz., $U = 1.07$ eV for the activation energy of the process and $\tau_0 = 2.4 \times 10^{-14}$ s for the pre-exponential factor of the relaxation time, are typical of phenomena caused by migration of point defects. The diffusion coefficient for point defects is estimated from the Einstein formula $D = \alpha \cdot a^2/\tau$, where the relaxation time $\tau = \tau_0 \exp(U/RT)$; for Nb the interplanar distance is $a = 3 \times 10^{-10}$ m, and the geometric coefficient for the bcc lattice is $\alpha = 1/24$.[5] The computed temperature dependence of the diffusion coefficient for point defects $D(T) = 2 \times 10^{-7} \exp(1.07\,\mathrm{eV}/RT)$ m^2/s is in satisfactory agreement with the diffusion equation for oxygen in niobium as calculated from a computer optimization of the aftereffect curves.[6] As is well known, when dissolved atoms undergo diffusion under stress in a bcc lattice (Snoek relaxation), their concentration c can be estimated from the simple formula

$$c(\mathrm{wt}\,\%) = K \cdot Q_p^{-1},$$

where Q_p^{-1} is the height of the internal friction peak after the background is subtracted off, and K is a coefficient of proportionality that is usually taken as ~ 1. This expression gives the following value of the oxygen concentration by weight for the composite fibers: 0.013 wt% for (Cu–8%Sn)–30%Nb and 0.035 wt% for (Cu–13%Sn)–30%Nb. According to manufacturer's certificate, the amounts of dissolved oxygen in the initial components, in the form of a vacuum-melted niobium sheet and rod niobium, are 0.01 and 0.02 wt%, respectively, i.e., the estimates of the oxygen content in the niobium are quite close to the manufacturer's data.

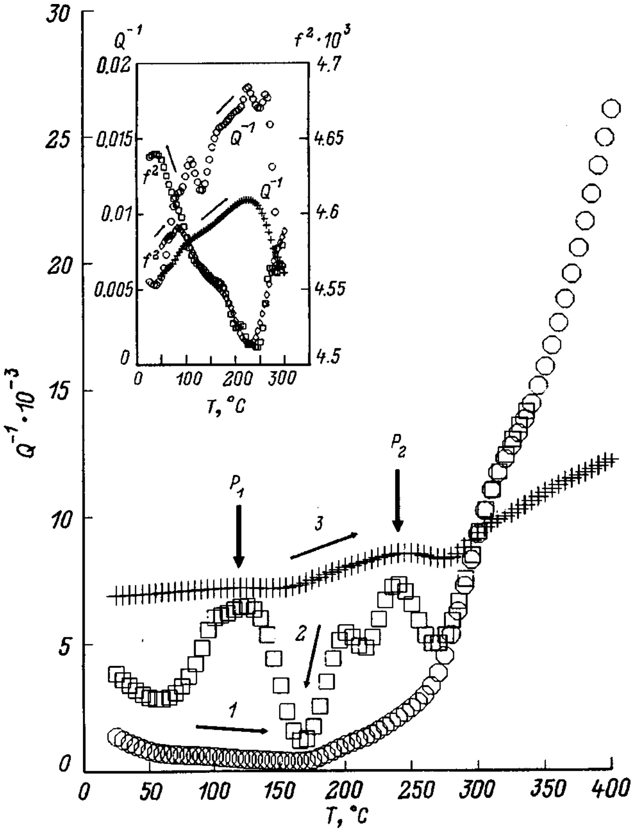

FIG. 2. Temperature dependence of internal friction of wires of diameter 0.5 mm and compositions (Cu–13%Sn)–30%Nb and Cu–25%Nb: *1* — heating curve for (Cu–13%Sn)–30%Nb, *2* — cooling curve for (Cu–13%Sn)–30%Nb, *3* — heating curve for Cu–25%Nb. The inset shows internal friction curves and the squared frequency of bending-resonance vibrations of a (Cu–8%Sn)–30%Nb composite cooled from 550 °C.

One possible reason for the hysteretic behavior of $Q^{-1}(T)$ in composite materials could be losses that occur as the materials are thermoplastically strained. This strain is caused by thermal stresses that appear during the cooling segment of the cycle, which in turn are due to differences in the thermal expansion coefficients of the components.[7,8] The following expression can be used to estimate the internal stresses in the composites under nonequilibrium thermal conditions[1]:

$$|\sigma_\mathrm{m}| = V_\mathrm{f} E_\mathrm{f} E_\mathrm{m} \Delta\alpha \Delta T / V_\mathrm{m} E_\mathrm{m} + V_\mathrm{f} E_\mathrm{f},$$

$$|\sigma_\mathrm{f}| = V_\mathrm{m} E_\mathrm{m} E_\mathrm{f} \Delta\alpha \Delta T / V_\mathrm{m} E_\mathrm{m} + V_\mathrm{f} E_\mathrm{f},$$

where σ_m and σ_f are the stresses in the matrix and fibers respectively, V_m and V_f are the volume fractions of matrix and fiber material, E_m and E_f are their Young's moduli, $\Delta\alpha = (\alpha_\mathrm{m} - \alpha_\mathrm{f})$ is the difference in thermal expansion coefficients of the matrix and fibers, and ΔT is the temperature difference.

In the case of fiber composites, it is assumed that these stresses appear only along the fiber axes and are constant over the cross sections of the components. For the cooling interval 400–100 °C, estimates for the (Cu–13%Sn)–30%Nb composite give values of roughly 9 kg/mm^2 for the bronze matrix and 21 kg/mm^2 for the niobium fibers. When added to the applied stresses $(0.1-1\ \mathrm{kg/mm}^2)$ at which the internal friction measurements were made, the thermal stresses generated are capable of causing plastic flow of both components of the composite[9] as the samples are cooled from 400 to 100 °C, i.e., to the location of peak P_1.

Since the volume fraction of matrix material $V_\mathrm{m}\cdot 100\% = 70\%$, it is natural to assume that below 140 °C the microplastic deformation of the matrix gives the dominant contribution to the total internal friction of the cooled composite. In this approximation, we can estimate the temperature dependence of the relative yield point $\sigma_\mathrm{y,m}(T)/\sigma_\mathrm{y,m}(20\,^\circ\mathrm{C})$ of the matrix in the temperature range where the plasticity peak P_1 exists. In a theoretical model that treats the dissipation of elastic energy as a function of the probability of appearance of plastic deformation in a microvolume,[10] the internal friction is proportional to the ratio of the power-law dependence of the stress in the material to that of its yield point:

$$Q^{-1} \sim \sigma^{n-2}_\mathrm{av,m}/\sigma^{n}_\mathrm{y,m},$$

where n is a constant that depends on the variance of the distribution function of the stress over microvolumes; for the estimative calculations we choose this to equal 3. Here $\sigma_\mathrm{av,m}$ is the average stress in the matrix at the surface of the sample being measured, and $\sigma_\mathrm{y,m}$ is the yield point of the matrix.

The statistical mechanism for the onset of microplasticity will make the most important contribution to the internal friction in the case of composite materials whose components differ greatly in their mechanical properties.[11]

To first approximation, we assume that the average $\sigma_\mathrm{av,m}$ is the same in order of magnitude as the thermal stresses in the matrix:

$$\sigma_\mathrm{av,m} \simeq V_\mathrm{f}E_\mathrm{f}E_\mathrm{m}\Delta\alpha\Delta T/V_\mathrm{m}E_\mathrm{m}+V_\mathrm{f}E_\mathrm{f}.$$

From this expression, knowing the thermal cycling parameters and using the previous formula for Q^{-1}, we obtain the desired normalized yield point $\sigma_\mathrm{y,m}(T)/\sigma_\mathrm{y,m}(20\,^\circ\mathrm{C})$. The results of the corresponding calculations are shown in Fig. 3. The curves shown in this figure imply that the recorded level of internal friction in the temperature range 20–140 °C corresponds to the calculated and experimentally observed changes in yield point of the matrix to within approximately 20%.

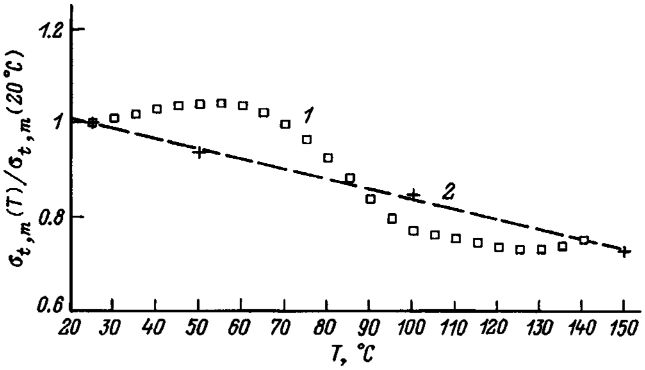

FIG. 3. Temperature dependence of the normalized yield points $\sigma_\mathrm{y,m}(T)/\sigma_\mathrm{y,m}(20\,^\circ\mathrm{C})$ of tin bronzes with comparable concentrations of Sn: *1* — calculated yield point of (Cu–13%Sn)–30%Nb bronze, *2* — experimental yield point of BR 014 bronze (See Ref. 4).

CONCLUSION

In the course of these studies of temperature-dependent internal friction in directed-crystallization (Cu–Sn)–Nb alloys we have discovered a temperature (∼175 °C) at which the internal friction is a minimum, which is correlated with the temperature at which the ''plasticity dip'' occurs in tin bronzes. We have also observed Snoek relaxation in niobium fibers due to oxygen dissolved in the niobium. The method of calculation we used here for the residual thermal stresses highlights the considerable potential of the internal friction method for reliably estimating the magnitude of the yield point of the matrix material. If the theory of microplasticity-induced acoustic losses developed here can be adapted to directed-crystallization structures, it may be possible to calculate quantitatively the temperature distribution of the yield points of the components of ''composite'' materials.

[1] K. I. Portnoĭ, B. N. Babich, and I. A. Svetlov, *Nickel-Based Composite Materials* [in Russian] (Metallurgiya, Moscow, 1979).

[2] S. A. Golovin and A. Pushkar, *Microplasticity and Metal Fatigue* [in Russian] (Metallurgiya, Moscow, 1980).

[3] M. Suenaga and W. Jansen, Appl. Phys. Lett. **43**, 791 (1983).

[4] I. M. Nikol'skaya and E. S. Shpichinetskiĭ, *Metallurgy and Processing of Nonferrous Metals and Alloys* [in Russian] (Metallurgiya, Moscow, 1968), pp. 7–10.

[5] M. A. Krishtall and S. A. Golivin, *Internal Friction and Structure of Metals* [in Russian] (Metallurgiya, Moscow, 1976), 475 pp.

[6] F. J. M. Baratto and R. E. Reed-Hill, Scr. Metall. **11**, 709 (1977).

[7] V. S. Postnikov, S. A. Ammer, and A. N. Kachevskiĭ, Pis'ma Zh. Tekh. Fiz. **5**, 560 (1979) [Sov. Tech. Phys. Lett. **5**, 229 (1979)].

[8] V. M. Arzhavitin, B. I. Shapoval, V. Ya., Sverdlov, and A. S. Tortika, Vopr. At. Nauki Tekh. Ser. Yad.-Fiz. Issled. Teor. Éksp. (Kharkov), **3**(3), pp. 66-68 (1989).

[9] L. V. Tikhonov, V. A. Kononenko *et al.*, *Structure and Properties of Metals and Alloys (Mechanical Properties of Metals and Alloys)* [in Russian] (Nakova Dumka, Kiev, 1986).

[10] V. T. Troshchenko, Fiz. Tverd. Tela (Leningrad) **2**, 1060 (1960) [Sov. Phys. Solid State **2**, 958 (1960)].

[11] I. I. Renne and I. N. Yurkin, *Internal Friction and Dislocation Structure of Metals* [in Russian] (Tula, 1990), pp. 79–82.

Translated by Frank J. Crowne

Effect of cubic magnetic anisotropy on angular dependences of the resonance field in (111)-oriented films

A. M. Zyuzin and V. V. Radaĭkin

Mordvinian State University, 430000 Saransk, Russia

(Submitted April 2, 1997; resubmitted August 11, 1997)
Zh. Tekh. Fiz. **68**, 118–120 (November 1998)

The angular dependences of the ferromagnetic resonance (FMR) field in (111)-oriented films are analyzed with the use of resonance relations and the conditions for equilibrium orientation of the magnetization. Based on the results obtained, an FMR method is proposed for determining the sign of the cubic anisotropy and the position of the crystallographic axes. [S1063-7842(98)02411-8]

In Ref. 1, Medved' *et al.* showed that cubic anisotropy has a negligible effect on the azimuthal dependence of the ferromagnetic resonance field $H_r(\varphi_H)$ of a film with the (111) orientation. In Refs. 2 and 3 we showed that the cubic anisotropy has its maximum influence when $H_r(\varphi_H)$ is measured in a geometry in which the angle θ_M between the normal to the film and the magnetization vector $\mathbf{M}$ is equal or close to 60°. In this case $\mathbf{M}$ passes close to both the $\langle 100\rangle$ and $\langle 111\rangle$ axes, which are either hard or easy axes, depending on the sign of the cubic anisotropy energy, and which lie in $\{100\}$ planes that are perpendicular to the film plane (111). As a result, the experimental function $H_r(\varphi_H)$ exhibits successive maxima and minima over each 60° sweep (Fig. 1). By recording the function $H_r(\varphi_H)$ in this way, we can determine the position of the $\{110\}$ planes.

The goal of this paper is to study the effect of cubic anisotropy on the polar dependence of the resonance field $H_r(\theta_H)$ in a $\{110\}$ plane perpendicular to the film plane (111). Calculation of the functions $H_r(\theta_H)$ (where θ_H is the angle between the field $\mathbf{H}$ and the normal to the film) were done with the help of the resonance relation[4]

$$\left(\frac{\omega}{\gamma}\right)^2=[H\cos(\theta_H-\theta_M)+H_k^{\mathrm{eff}}\cos^2\theta_M-H_{k1}a(\theta_M)]\times[H\cos(\theta_H-\theta_M)+H_k^{\mathrm{eff}}\cos 2\theta_M-H_{k1}b(\theta_M)],\qquad(1)$$

where ω is the angular frequency of the microwave field, γ is the gyromagnetic ratio, $H_k^{\mathrm{eff}}=H_{ku}-4\pi M_s$ is the effective uniaxial anisotropy field, $H_{k1}=2K_1/M$ is the cubic anisotropy field, H is the magnitude of the external magnetic field applied in the $\{110\}$ plane at resonance, and

$$a(\theta_M)=(3-16x-3y)/16,\quad b(\theta_M)=-(x+3y)/4,$$

$$x=(2^{3/2}\sin 2\theta_M-\cos 2\theta_M)/3,$$

$$y=-(2^{5/2}\sin 4\theta_M+7\cos 4\theta_M)/9.$$

The positive angles θ_H and θ_M are measured in the (110) plane from the normal to the film [111] to the crystallographic direction [001] along the shortest path. The angle between the magnetization vector $\mathbf{M}$ and the normal to the film θ_M corresponding to a fixed value of θ_H is determined from the condition of equilibrium orientation of the magnetization:[4,5]

$$2H\sin(\theta_H-\theta_M)=H_k^{\mathrm{eff}}\sin 2\theta_M+H_{k1}\times\left[\frac{1}{12}\sin 2\theta_M+\frac{7}{24}\sin 4\theta_M+\sqrt{2}\sin^2\theta_M-\frac{4\sqrt{2}}{3}\sin^4\theta_M\right].\qquad(2)$$

The polar dependence $H_r(\theta_H)$ is calculated in the following way. First of all, the resonance fields $H_\perp$ for the perpendicular orientation and $H_\parallel$ for the parallel, whose value are known from experiment, are substituted into the expressions

$$\frac{\omega}{\gamma}=H_\perp+H_k^{\mathrm{eff}}-\frac{2}{3}H_{k1},\qquad(3)$$

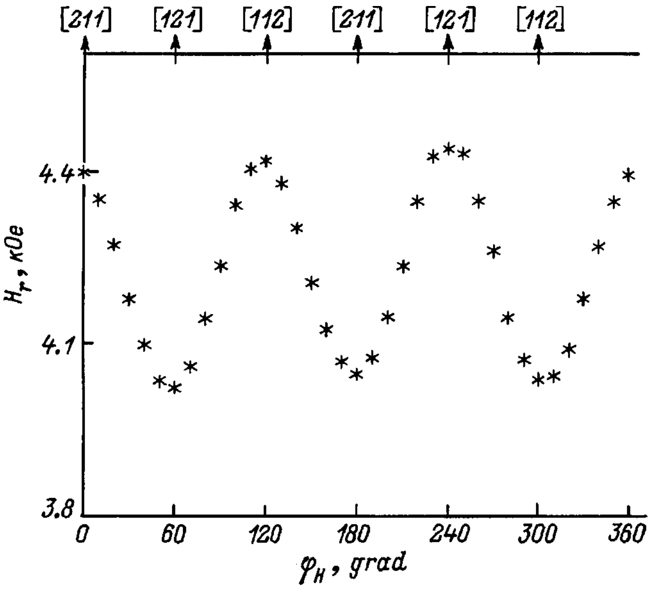

FIG. 1. Experimental azimuthal dependence of the resonance field $H_r(\varphi_H)$ for θ_H corresponding to $\theta_M=60°$ (a sample with the composition $Eu_{1.4}Lu_{1.6}Fe_5O_{12}$).

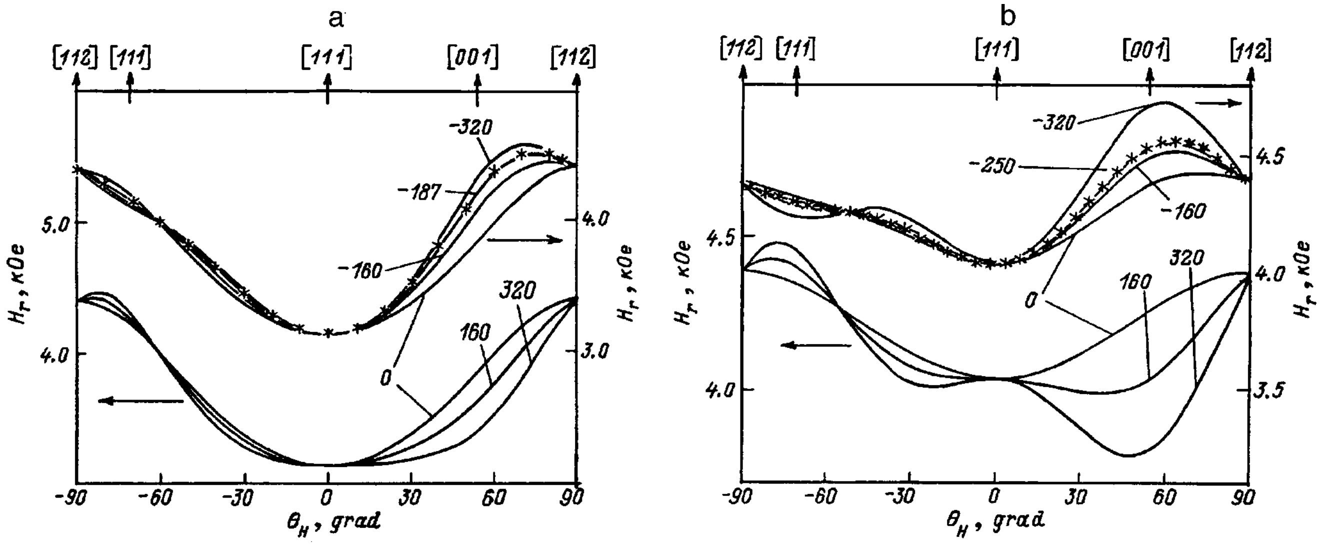

FIG. 2. Polar functions $H_r(\theta_H)$. The curves are calculated, the crosses experiment for samples with the compositions $Eu_{1.4}Lu_{1.6}Fe_5O_{12}$ (a) and $Eu_1Tm_2Fe_5O_{12}$ (b). The numbers on the curves are the values of H_{k1} (Oe).

and

$$\left(\frac{\omega}{\gamma}\right)^2 = H_\parallel\left(H_\parallel - H_k^{\mathrm{eff}} - \frac{1}{2}H_{k1}\right). \tag{4}$$

From these we can calculate γ and H_k^{eff}. Starting with H_{k1} set equal to zero, we then calculate the function $H_r(\theta_H)$ for this choice of γ, H_k^{eff} and H_{k1}, using (1) and (2).Then, choosing another value of H_{k1} and the same $H_\perp$, $H_\parallel$, we calculate γ and H_k^{eff} and again evaluate the function $H_r(\theta_H)$. Thus we obtain a family of polar functions corresponding to fixed values of $H_\perp$ and $H_\parallel$.

The family of curves $H_r(\theta_H)$ obtained for various H_{k1} are shown in Figs. 2a and 2b, along with experimental functions for single-crystal films of iron garnets with the compositions $Eu_{1.4}Lu_{1.6}Fe_5O_{12}$ ($H_k^{\mathrm{eff}} = 765$ Oe, $H_{k1} = -250$ Oe, $\gamma = 1.44 \times 10^7$ Oe^{-1}s^{-1}) (a) and $Eu_1Tm_2Fe_5O_{12}$ ($H_k^{\mathrm{eff}} = 182$ Oe, $H_{k1} = -187$ Oe, $\gamma = 1.35 \times 10^7$ Oe^{-1}s^{-1}) (b). It is seen that the experimental $H_r(\theta_H)$ curves are in good agreement with the calculated curves for suitable values of H_k^{eff}, H_{k1}, and γ. The method used to determine the values of H_{k1} in the films was described in Ref. 3. The measurements were made at room temperature and at a microwave field frequency of 9.34 GHz. The figures show that the $H_r(\theta_H)$ curves obtained for various values of H_{k1} have a point of intersection with the calculated $H_r(\theta_H)$ curve for $H_{k1} = 0$ (starting curve) in the range $(-90^\circ < \theta_H < 0)$. This intersection point is common to all families of curves and is located near $\theta_M = 60^\circ$ in the quadrant containing the $\langle 111 \rangle$ crystallographic axis. In the adjacent quadrant, which contains the $\langle 100 \rangle$ axis, there is no such point. Depending on the sign of H_{k1}, the $H_r(\theta_H)$ curves are located in this quadrant $(0 < \theta_H < 90^\circ)$ either above (for $H_{k1} < 0$) or below (for $H_{k1} > 0$) the curve for $H_{k1} = 0$. Moreover, it is not difficult to see that there is also a characteristic ordering in the positions of these curves in the first quadrant $(-90^\circ < \theta_H < 0)$. Note that the polar function $H_r(\theta_H)$ also possesses analogous properties when the values of the effective uniaxial anisotropy field H_k^{eff} are negative.

Thus, by analyzing the experimental azimuthal dependence of the resonance field $H_r(\varphi_H)$ (obtained, e.g., for θ_H corresponding to $\theta_M = 60^\circ$) we can avoid resorting to x-ray methods to determine the position of the $\{110\}$ crystallographic planes perpendicular to the film plane (111).

Using the characteristic indicators of the shape of the experimental polar dependence, we can compare it with curves calculated for the same values of resonance field in the perpendicular and parallel orientations and with $H_{k1} = 0$ and thereby determine the sign of the cubic anisotropy constant and the position of the $\langle 100 \rangle$ and $\langle 111 \rangle$ axes in the $\{110\}$ planes mentioned above.

Thus, using $H_\perp$ and $H_\parallel$ and the minimum and maximum values of the azimuthal dependence of the resonance field, we can calculate the values of the cubic anisotropy field, the uniaxial anisotropy field, and the gyromagnetic ratio.

[1] A. V. Medved', N. P. Nikitin, and L. M. Filimonova, Zh. Éksp. Teor. Fiz., 1557 (1985).
[2] A. M. Zyuzin and Al. M. Zyuzin, Fiz. Tverd. Tela (Leningrad) **29**, 3128 (1987) [Sov. Phys. Solid State **29**, 1795 (1987)].
[3] A. M. Zyuzin, V. V. Radaĭkin, and A. G. Bazhanov, Zh. Tekh. Fiz. **67**(2), 35 (1997) [Tech. Phys. Tech. Phys. **42**, 155 (1997).
[4] J. Makino and Y. Hidaka, Mat. Res. Bull. Vol. 16, N 8, 957 (1981).
[5] A. G. Gurevich, *Magnetic Resonance in Ferrites and Antiferromagnets* [in Russian] (Nauka, Moscow, 1973), 591 pp.

Translated by Frank J. Crowne

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Reciprocity relations for the nonlinear conductivity of fractal resistors

A. A. Snarskiĭ and S. I. Buda

National Technical University of Ukraine, 25205 Kiev, Ukraine

(Submitted July 15, 1997)
Zh. Tekh. Fiz. **68**, 121–124 (November 1998)

Reciprocity relations are derived for the conductivity of finite-size samples of inhomogeneous weakly nonlinear two-dimensional media. © *1998 American Institute of Physics.* [S1063-7842(98)02511-2]

INTERACTION

The nonlinear properties of highly inhomogeneous composites have always been a focus of interest (see, for example, Ref. 1 and the papers cited therein). One particular reason for this interest is that the distributions of fields and currents in these media are highly nonuniform, which forces us to take into account deviations from Ohm's law.

For macroscopically inhomogeneous media, to first approximation with respect to the nonlinearity we replace the linear Ohm's law by

$$\mathbf{j}(\mathbf{r})=\sigma(\mathbf{r})\mathbf{E}(\mathbf{r})+\chi(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2\mathbf{E}(\mathbf{r}), \tag{1}$$

where $\mathbf{j}(\mathbf{r})$ and $\mathbf{E}(\mathbf{r})$ are the electrical current density and electric field, and $\sigma(\mathbf{r})$ and $\chi(\mathbf{r})$ are the linear and nonlinear electrical conductivities.

In two-phase media, which will be topic of discussion below, the electrical conductivities have values σ_1, χ_1 and σ_2, χ_2 in the first and second phases, respectively. We will discuss the case of strong nonuniformity, where $h=\sigma_2/\sigma_1\ll 1$, and weak nonlinearity, where $\chi(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2\ll\sigma(\mathbf{r})$.

The most important characteristics of randomly inhomogeneous media are their effective kinetic coefficients σ_e and χ_e, which are defined so as to relate the field and current throughout the volume:

$$\langle\mathbf{j}(\mathbf{r})\rangle=\sigma_e\langle\mathbf{E}(\mathbf{r})\rangle+\chi_e\langle\mathbf{E}(\mathbf{r})\rangle^2\langle\mathbf{E}(\mathbf{r})\rangle,\langle\dots\rangle$$
$$=V^{-1}\int\dots dV. \tag{2}$$

Over a characteristic averaging dimension $L\propto V^{1/3}$ much larger than the correlation length ξ, a self-averaging of these system parameters takes place. The effective conductivity depends on the concentration. For example, in randomly inhomogeneous media near the percolation threshold[2]

$$\sigma_e\approx\sigma_1\tau^t,\quad \tau=(p-p_c)/p_c,\quad \tau>0, \tag{3}$$

where p is the concentration of the phase with conductivity σ_1, p_c is the percolation threshold, and t is the critical exponent of the conductivity.

Expression (3) is asymptotically correct when $\tau\to 0$, in which limit we can neglect percolation of the current through the phase with conductivity σ_2. Analogous universal behavior of the conductivity is also obtained below the percolation threshold, i.e., for $\tau<0$. Universal relations of type (3) have also been established for χ_e, but obviously with different critical exponents.[3] The question of what should be taken as the representative size ξ for such media, in the terminology of I. M. Lifshits,[4] requires a special discussion in each individual case. For two-phase percolation media we have

$$\xi\propto a_0|\tau|^{-\nu}, \tag{4}$$

where a_0 is the minimum size in the system (the "grain" size of the inhomogeneity), and ν is the critical exponent for the correlation length.[5]

For samples with dimensions $L<\xi$ the system is mesoscopic, and the measured characteristics fluctuate from realization to realization. In this case, the well-defined physical quantities are averages over realizations. Because in percolation systems the effective conductivities $\{\sigma_e\}$ and $\{\chi_e\}$ averaged over realizations are power-law functions of the system size,[5] such systems are referred to as fractal. When the contribution of one of the phases to the conductivity of the entire system can be neglected, e.g., for $\tau>0$, $\sigma_2=0$, there exists a "recipe"[5] for determining how the parameters averaged over realizations depend on the size L of the system. Setting $L=\xi$, solving for τ in Eq. (4), and substituting the result $\tau=(L/a_0)^{-1/\nu}$ into Eq. (3) for σ_e, we obtain

$$\{\sigma_e\}\approx\sigma_1(L/a_0)^{-1/\nu},\quad \tau>0, \tag{5}$$

and proceed analogously for the other effective coefficients averaged over realizations.

When $h=\sigma_2/\sigma_1\neq 0$ this simple "recipe" no longer applies.[5] For inhomogeneous media that are far from the percolation threshold, where there is no universal dependence of type (3), and for media whose special structure prevents us from defining the concept of proximity to the percolation threshold at all, universal functions of type (5) cannot be found. Nevertheless, as we show below, for a certain class of two-dimensional media (including percolation media) it is possible to derive exact reciprocity relations for averages over realizations. These reciprocity relations interrelate certain combinations of the conductivity coefficients averaged over realizations and are generalizations of the reciprocity relations for the linear case in media that are on average geometrically equivalent with respect to the arrangement of the phases; those relations were first established for $L\gg\xi$ in Refs. 7 and 8.

1063-7842/98/43(11)/4/$15.00

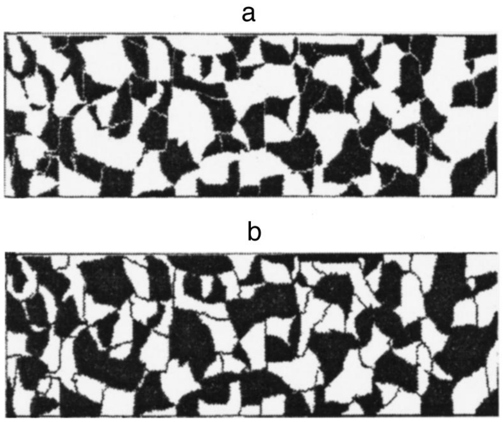

FIG. 1. Inhomogeneous finite-size sample (a) and the dual sample obtained from the first by exchanging phases $\sigma_1 \leftrightarrow \sigma_2$, $X_1 \leftrightarrow X_2$ (b).

DERIVING THE RELATIONS BETWEEN FUNCTIONS

Let us consider an inhomogeneous two-dimensional sample of size $L_\parallel \times L_\perp$ and unit thickness. The currents and voltage drops $I_\parallel$, $U_\parallel$ and $I_\perp$, $U_\perp$ along and transverse to the long side, respectively, are related according to Eq. (2) by the linear and nonlinear conductances

$$I_\parallel = G_\parallel U_\parallel + X_\parallel U_\parallel^3, \quad I_\perp = G_\perp U_\perp + X_\perp U_\perp^3. \tag{6}$$

Here we assume that in the first case ideal contacts (with zero resistance) are applied to the vertical faces of the sample, while in the second case the contacts are applied to the horizontal faces. According to Ref. 8, we can find a corresponding sample that is dual to any sample of an inhomogeneous medium:

$$j = \Lambda \tilde{P}_{\pi/2} \tilde{E}, \quad E = \Lambda^{-1} \tilde{P}_{\pi/2} \tilde{j}, \tag{7}$$

where $\Lambda = \sqrt{\sigma_1 \sigma_2}$, and $\tilde{P}_{\pi/2}$ is the operator of rotation in the plane of the medium by an angle $\pi/2$.

In what follows we will assume that the following relation holds for the local coefficients:

$$\chi_2 / \chi_1 = -(\sigma_2 / \sigma_1)^2. \tag{8}$$

In this case it is easy to show that substituting Eq. (6) into Eq. (1) leads to expressions that describe, with accuracy up to terms cubic in the field, a medium with the same local conductivity law

$$\langle \tilde{\mathbf{j}}(\mathbf{r}) \rangle = \tilde{\sigma}_e \langle \tilde{\mathbf{E}}(\mathbf{r}) \rangle + \tilde{\chi}_e \langle \tilde{\mathbf{E}}(\mathbf{r}) \rangle^2 \langle \tilde{\mathbf{E}}(\mathbf{r}) \rangle \tag{9}$$

but with mutually exchanged phases: if $\sigma(\mathbf{r} \in 0_1) = \sigma_1$, $\sigma(\mathbf{r} \in 0_2) = \sigma_2$, then $\sigma(\mathbf{r} \in 0_1) = \sigma_2$, $\sigma(\mathbf{r} \in 0_2) = \sigma_1$ in the new medium, and analogously for $\chi(\mathbf{r})$. That is, if we color the first phase black and the second white, the dual sample will be the ‘‘negative’’ of the first, and the boundary conditions, determined by the set of contacts (Fig. 1b), will be modified accordingly. In this case the same equations $(\nabla \cdot \tilde{\mathbf{j}} = 0$, $\nabla \times \tilde{\mathbf{E}} = 0)$ will hold in the dual medium as hold in the initial medium.

After averaging (7), we obtain for the total currents and voltage drops

$$I_{\parallel,\perp} = \Lambda U_{\perp,\parallel}, \quad U_{\parallel,\perp} = \Lambda^{-1} I_{\perp,\parallel}. \tag{10}$$

Applying these transformations to Eq. (5), to accuracy up to cubic terms we are led to

$$\tilde{I}_{\parallel,\perp} = \tilde{G}_{\parallel,\perp} \tilde{U}_{\parallel,\perp} + \tilde{X}_{\parallel,\perp} \tilde{U}_{\parallel,\perp}^3, \tag{11}$$

where

$$\tilde{G}_{\parallel,\perp} = \Lambda^2 / G_{\perp,\parallel}, \quad \tilde{X}_{\parallel,\perp} = -X_{\perp,\parallel} \Lambda^4 / G_{\perp,\parallel}^4. \tag{12}$$

From these two relations all the results of this paper will follow.

RECIPROCITY RELATIONS FOR RANDOMLY INHOMOGENEOUS MEDIA IN THE FRACTAL REGIME

If the sample consists of a portion of randomly inhomogeneous medium and $L_\parallel = L_\perp \gg \xi$, then $G_\parallel = G_\perp = \sigma_e$, $\tilde{G}_\parallel = \tilde{G}_\perp = \tilde{\sigma}_e$ (recall that the sample has unit thickness). Then the Dykhne reciprocity relation[8] $\sigma_e \tilde{\sigma}_e = \sigma_1 \sigma_2$ follows from Eq. (12). Because $\tilde{\sigma}_e(p) = \sigma_e(1-p)$ for randomly inhomogeneous media, the latter can be written as $\sigma_e(p) \sigma_e(1-p) = \sigma_1 \sigma_2$.

When $L < \xi$ it is necessary to average over realizations. There are several ways to transform from Eq. (12) to an average over realizations. In particular, from the first relation in Eq. (12) it follows[9] that

$$\{\tilde{G}_{\perp,\parallel}\} / \{R_{\parallel,\perp}\} = \sigma_1 \sigma_2, \tag{13}$$

where $R_{\parallel,\perp} = 1/G_{\parallel,\perp}$ is the sample resistance.

For $L_\parallel = L_\perp$ and $p = p_c$ the samples are cut out from a medium at the percolation threshold, and each of them is naturally fractal; hence, the averages over realizations in the starting and dual media coincide. Then from Eq. (13) we obtain[7]

$$\{G\} / \{R\} = \sigma_1 \sigma_2, \quad L_\parallel = L_\perp, \quad p = p_c. \tag{14}$$

Yet another reciprocity relation can be obtained for the linear part of the conductance if we take the logarithm of Eq. (12) before averaging over realizations. Taking into account that $\{\ln \tilde{G}_\perp\} = \{\ln G_\parallel\}$ when $L_\parallel = L_\perp$ and $p = p_c$, we obtain

$$\{\ln G\} = \ln \sqrt{\sigma_1 \sigma_2}, \quad L_\parallel = L_\perp, \quad p = p_c. \tag{15}$$

Analogous but more complicated reciprocity relations can be written for the nonlinear part of the conductance. In particular, it follows from Eq. (12) that

$$\frac{\tilde{X}_\perp}{\tilde{G}_\perp^2} = -\frac{X_\parallel}{G_\parallel^2}, \quad \frac{\tilde{X}_\parallel}{\tilde{G}_\parallel^2} = -\frac{X_\perp}{G_\perp^2}, \tag{16}$$

which after averaging over realizations gives

$$\left\{ \frac{\tilde{X}_\perp}{\tilde{G}_\perp^2} \right\} = -\left\{ \frac{X_\parallel}{G_\parallel^2} \right\}, \quad \left\{ \frac{\tilde{X}_\parallel}{\tilde{G}_\parallel^2} \right\} = -\left\{ \frac{X_\perp}{G_\perp^2} \right\}. \tag{17}$$

We note that Eq. (17) for $L_\parallel = L_\perp$ and $p = p_c$ gives

$$\left\{ \frac{X}{G^2} \right\} = -\left\{ \frac{X}{G^2} \right\} \quad \text{and thus} \quad \left\{ \frac{X}{G^2} \right\} = 0. \tag{18}$$

We now consider the case where one of the dimensions of the sample is smaller than the correlation length, as before, while the other is larger, i.e., $L_\parallel \gg \xi$, $L_\perp \ll \xi$, or the case of a long strip. Despite the fact that one of the dimensions is smaller than the correlation length, it is not necessary to average over realizations of the randomly inhomogeneous medium. In fact, $\{G_\parallel\} = G_\parallel$, where $G_\parallel$ is the conductance of a given realization, and analogously for other conductances. Moreover, for a strip cut out from a medium at the percolation threshold we have $\tilde{G}_\parallel = G_\parallel$, $\tilde{X}_\parallel = X_\parallel, \ldots$, and we can obtain from Eq. (12)

$$G_\parallel G_\perp = \Lambda^2, \quad X_\parallel / X_\perp = -(G_\parallel / G_\perp)^2,$$

$$L_\perp \ll \xi, \quad L_\parallel \gg \xi. \tag{19}$$

Thus, despite the fact that the linear and nonlinear conductances are power-law functions of $L_\perp$, certain combinations of them, in particular Eqs. (19), do not depend on the dimension $L_\perp$.

RECIPROCITY RELATIONS FOR DETERMINISTIC STRUCTURES

Over distances smaller than the correlation length, the conductances of individual realizations of a randomly inhomogeneous medium are not of much interest in general. However, for certain media with deterministic inhomogeneity that possess certain symmetries, reciprocity relations of a given realization can be informative. Figure 2 shows inhomogeneous samples that possess symmetry with respect to rotation by $\pi/2$ and mutual exchange of phases. When this symmetry holds we have $G_\parallel = G_\perp$, $X_\parallel = X_\perp$, and the dual medium has the same conductance as the original one. From Eq. (12) it follows in this case that

$$G = \sqrt{\sigma_1 \sigma_2}, \quad X = 0. \tag{20}$$

Recall that in the case we are considering the local conductances satisfy condition (8). If, for example, we have $\chi_1 > 0$ in the first phase (a superlinear current–voltage (I–V) characteristic) and $\chi_2 < 0$ in the second (a sublinear I–V characteristic), then in the structures under study the superlinear behavior of one phase exactly compensates the sublinear behavior of the other, and a nonlinear local sample will have an I–V characteristic that is linear overall.

It is also interesting to consider the case of a long strip with deterministic inhomogeneity. Figure 3a shows a layerwise-inhomogeneous medium for which it is not difficult to compute the conductances along and transverse to the inhomogeneity. With accuracy up to small higher-order nonlinear terms we have

$$G_\parallel = \langle G(y) \rangle; \quad X_\parallel = \langle X(y) \rangle; \quad G_\perp = 1/\langle 1/G(y) \rangle;$$

$$X_\perp = \langle X(y)/(G(y))^4 \rangle / \langle 1/G(y) \rangle^4. \tag{21}$$

On the other hand, mutually exchanging phases in the plane-layered sample under study does not change the conductances $\tilde{G}_\parallel = G_\parallel, \ldots$, and according to Eq. (12)

$$G_\parallel G_\perp = \sigma_1 \sigma_2, \quad X_\parallel / X_\perp = -(G_\parallel / G_\perp)^2. \tag{22}$$

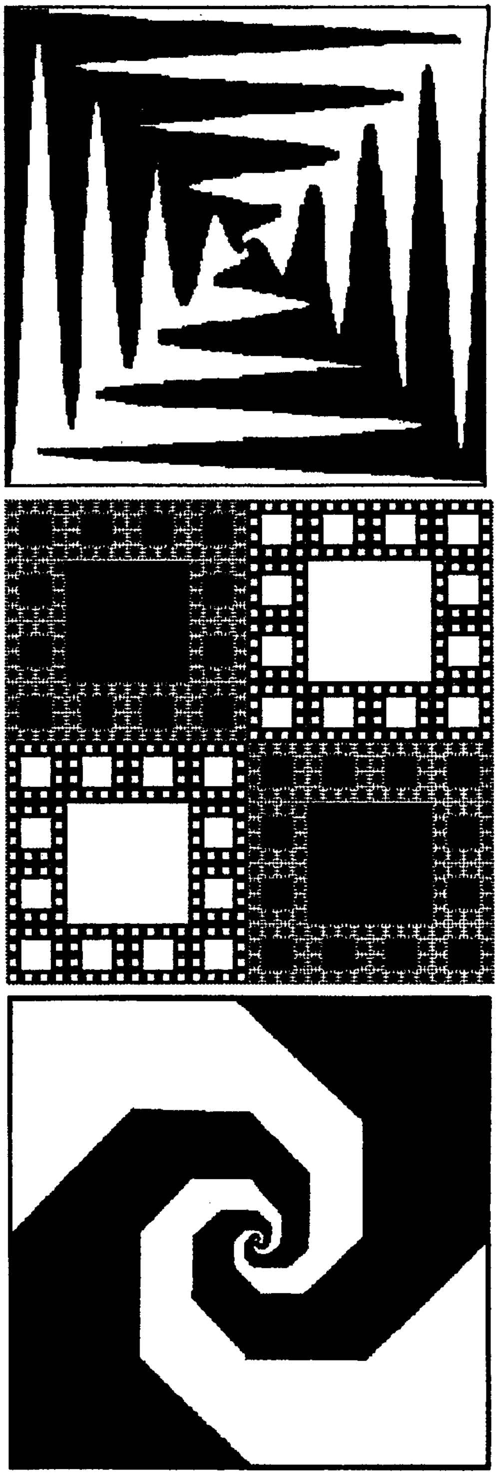

FIG. 2. Two-phase dual structures. Under an exchange of phases and rotation by $\pi/2$ the resistance of the sample remains unchanged.

It is easy to show that including condition (8) in Eq. (21) and substituting the latter into Eq. (22) leads to an identity.

One more example of an inhomogeneous strip is shown in Fig. 3b — a two-phase film with a ''rough'' boundary between phases. If the boundary has symmetry such that mutual exchange of the phases does not lead to a change in conductances, then for this sample the reciprocity rule (22) holds. For example, the boundary between phases could have the form of a sine for an arbitrary value of the dimension $L_\parallel$

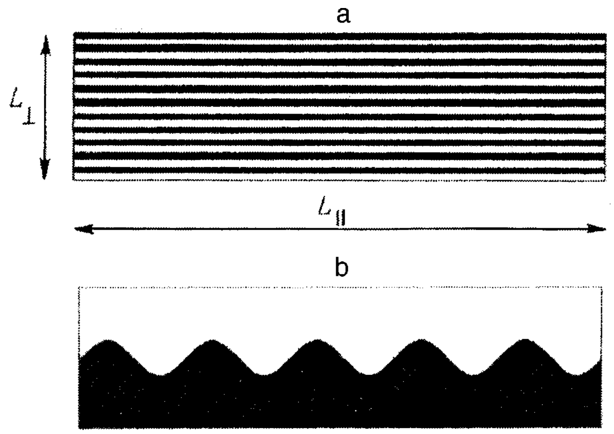

FIG. 3. Plane-layered (a) dual structure with a sinusoidal boundary between phases (b).

such that this length equals an integer number of periods. The shape of the boundary can also be random. Boundaries of this kind were first introduced in investigations of the reflection of electromagnetic waves from surfaces with shallow profiles. In Ref. 9 the authors obtained relations analogous to the first in Eq. (22) for the effective dielectric constant.

This work was partially supported by The Russian Foundation for Basic Research, Project No. 97-02-16923a.

[1] K. W. Yu and P. M. Hui, Phys. Rev. B **50**, 13 327 (1994); X. Zhang and D. Stroud, Phys. Rev. B **49**, 944 (1994).
[2] A. L. Efros and B. I. Shklovskii, Phys. Status Solidi B **76**, 475 (1976).
[3] D. Stroud and P. M. Hui, Phys. Rev. B **37**, 8919 (1988); R. Rammal, C. Tannous, and A.-M. S. Tremblay, Phys. Rev. A **31**, 2662 (1985); A. E. Morozovskiĭ and A. A. Snarskiĭ, Zh. Eksp. Teor. Fiz. **95**, 1844 (1989) [Sov. Phys. JETP **68**, 1066 (1989)]; O. Levy and D. J. Bergman, Phys. Rev. B **50**, 3652 (1994).
[4] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems* (Wiley, New York, 1998; Nauka, Moscow, 1982).
[5] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor & Francis, London, 1992).
[6] A. E. Morozovskiĭ and A. A. Snarskiĭ, Zh. Éksp. Teor. Fiz. **109**, 674 (1996) [Sov. Phys. JETP **82**, 361 (1996)].
[7] J. B. Keller, J. Appl. Phys. **34**, 991 (1963); J. Math. Phys. **5**, 548 (1964).
[8] A. A. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
[9] R. Z. Vitlina and A. A. Dykhne, Zh. Éksp. Teor. Fiz. **99**, 1758 (1991) [Sov. Phys. JETP **72**, 983 (1991)].

Translated by Frank J. Crowne

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Effect of static and dynamic compression on the healing of pores in copper

A. I. Petrov, M. V. Razuvaeva, A. V. Sinani, and V. V. Nikitin

A. I. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

(Submitted June 28, 1997)
Zh. Tekh. Fiz. **68**, 125–127 (November 1998)

The distinctive features of the healing of pores in copper (with an initial porosity of 12%) by various types of static and shock-wave-induced stresses are discussed. It is shown for static loading that the process of healing is determined by the maximum shear stress at the surface of a pore and the character of the distribution of shear stresses near the pore. The observed increase in the efficiency of healing of the porosity when dynamic methods of compression are used as compared to static methods is attributed to an increase in temperature due to local heating near the pore. © *1998 American Institute of Physics.* [S1063-7842(98)02611-7]

It is known[1] that at ordinary temperatures the healing of pores in crystalline materials under stress takes place via a dislocation mechanism. According to Ref. 1 shear stresses arise near a pore under pressure. When a stress is reached that exceeds the critical stress for triggering a Frank–Read source near the pore, a dislocation loop forms and the boundary of the pore is displaced by the magnitude of the Burgers vector. The degree of healing of the pore is determined by the number of dislocation loops emitted by this source, which depends on the mobility of the dislocations as well as the magnitude of the stress and the time within which it acts. We may therefore expect that the process of healing will be determined by the magnitude of the maximum shear stress $\tau_{\max}$ at the pore surface, the character of the stress (τ) distribution near the pore, and the time the stress acts. For this reason it is interesting to study the kinetics of pore healing in the presence of various kinds of dynamic and static compressions that differ in the parameters listed above. This will be the topic of the present paper.

The system under study was porous copper. Samples in the form of disks 18 mm in diameter with thickness $h = 1$ mm were obtained by evaporating the zinc from brass (Cu + 26%Zn) in vacuum at a temperature of 800 °C for 8 hours. The initial degree of porosity $W_0=(\rho_0-\rho)/\rho_0$, where ρ_0 is the tabulated density of copper and ρ is the density of the evaporated samples, was about 12%. The average size of a pore was 25 μm. Static forces were applied under conditions of hydrostatic pressure, uniaxial compression, and uniaxial compression with a clamped lateral surface. Dynamic compression of the copper samples was implemented by shock-wave loading of a disk placed in a ring restraint in order to prevent transverse spreading of the material. The duration of the compression pulse was $6-8$ μs for various pressure amplitudes P in the plane of the shock wave. A detailed description of the methods for imposing shock loading was given in Ref. 2. The porosity W was monitored before and after the various loadings by measuring the density and using an optical microscope (Neophot-30).

In Fig. 1 we plot how W varies under static (*1*) and dynamic (*2*) compression. It is clear that regardless of experimental conditions the magnitude of the threshold pressure P^* does not depend on the kind of compression and is ≈ 100 MPa. This implies that the value of critical stress that causes the healing process to begin is also the same for static and dynamic compression. However, at pressures higher than threshold ($P>P^*$) shock compression leads to a higher degree of healing than static.

The observed increase in the efficiency for dynamic compression as compared to static can be attributed to an increase in plasticity due to the rise in temperature. It is known[3] that the passage of a shock wave causes various thermal effects; at internal boundaries, such as pores, microcumulation effects can arise[4] that are accompanied by considerable additional heating.

Another reason for the increased mobility of dislocations under dynamic loading could be local heating in slip bands due to adiabatic processes during the high-velocity motion of

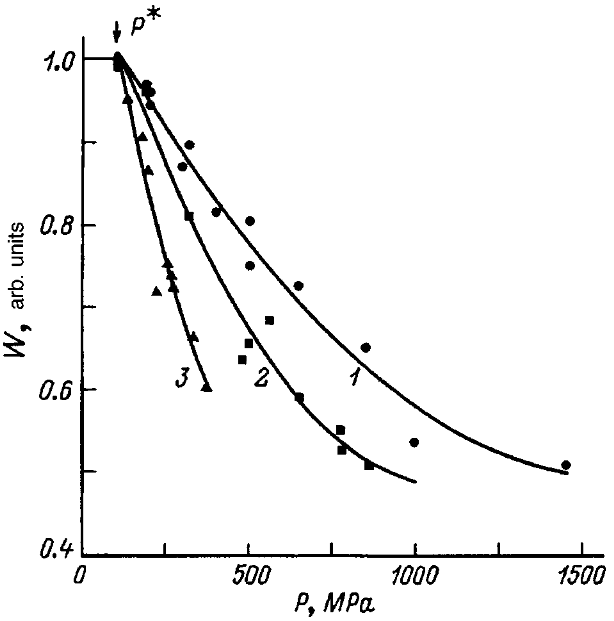

FIG. 1. Relative values of the porosity of copper after static (*1*), shock (*2*), and pulsed magnetic (*3*) compression. A value of 1.0 corresponds to $W_0 = 0.12$.

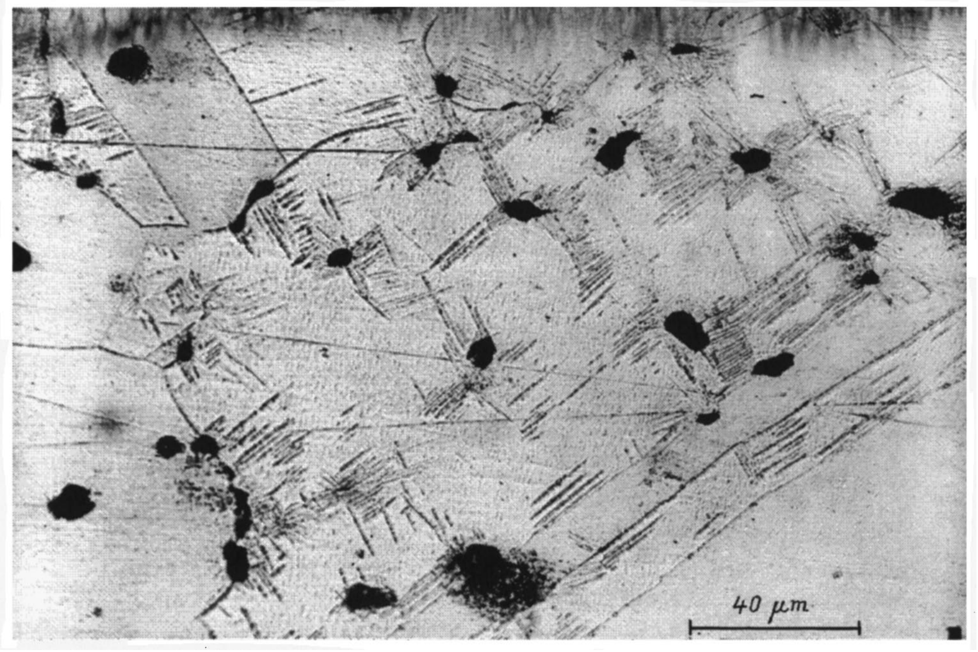

FIG. 2. Optical photograph of a section of a porous copper sample after the passage of a shock wave with $P = 2000$ MPa. Magnification 640×.

dislocations. A consequence of these processes is localization of the plastic deformation in shear bands or in adiabatic shear bands.[3] Metallographic analysis of sections of copper samples that had been subjected to dynamic compression show a high degree of localization of the plastic deformation near the pores. A decrease in the pore size is accompanied by the formation of shear bands near the pores (Fig. 2), the length of these bands increasing with increasing amplitude of the shock-wave compression. Under static compression these bands are not observed, which indicates that plastic deformation occurs homogeneously under these conditions.

Let us compare the effect of compression pulses of various durations on the porosity of copper. In Ref. 5 the authors studied the effects of pulsed magnetic pressures of different amplitudes with the same compression pulse duration $t \approx 2\ \mu$s. In Fig. 1 we have taken data from Ref. 5 to show the dependence of porosity on the magnitude of the magnetic pressure in the range from 0 to 350 MPa. A comparison of curves *2* and *3* shows that the degree of healing under dynamic loading is higher for shorter durations of the compression pulse. The results obtained contradict the notions of Ref. 1, according to which a decrease in the time the pressure acts should lead to a decrease in the healing effect.

We can assume that the primary cause of the increase in healing effect with decreasing pulse duration is due mainly to the steeper rise of the leading edge of the compression wave, since this leads to an increase in the rate of strain, enhancement of adiabatic effects, and local heating. Moreover, if the pore healing time is shortened, the stress field around the pore is less attenuated by rarefaction waves arriving from the surrounding pores. These waves can play a vital role in terminating the healing process.

Let us consider data on how the kind of stress state affects pore healing. Figure 3 shows the dependence of *W* on *P* for various modes of static compression that differ in respect to $\tau_{\max}$ and in the character of the decay of τ from the pore into the depths of the material. For hydrostatic compression (*I*) we have $\tau_{\max}=0.75P$, and the stress is short-ranged, falling off with increasing distance from the pore as $1/r^3$ (see Ref. 1), where r is the distance from its center. For uniaxial compression (*II*) and uniaxial compression with a clamped lateral surface (*III*) long-range stresses appear, in which case $\tau_{\max}$ equals $0.7P$ and $0.3P$, respectively.[6] Consequently, the differences we see in Fig. 3 are connected both with the magnitude of $\tau_{\max}$ and with the distribution of stresses near the pore. In fact a comparison of *I* and *II* shows that the value

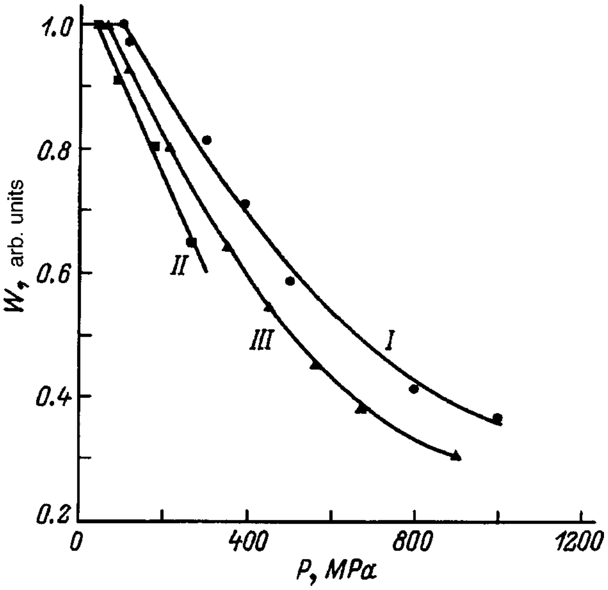

FIG. 3. Relative mean porosity of copper after hydrostatic compression (*I*), uniaxial compression with a clamped lateral surface (*III*), and uniaxial compression with a free surface (*II*).

of $\tau_{\max}$ under these conditions is the same but that the shear stresses are different, being long-ranged in the latter case. Because of this, the number of dislocations emitted by the pores in the compressed disk is determined only by the presence of obstacles to the motion of dislocations in the material, whereas under hydrostatic pressure it is determined by the opposing stresses from the dislocations emitted by the pores.[1] Uniaxial compression applied in different manners leads to different degrees of healing, which are determined by the value of $\tau_{\max}$ (curves *II* and *III* in Fig. 3).

Differences in the stress state for various methods of compression determine not only the different degree of healing but also the magnitude of the threshold pressure P^* (Fig. 3) and also the strain $\varepsilon = \Delta h/h$ that accompanies the process of healing. We have established that for hydrostatic compression and compression with a clamped lateral surface the well-known relations $\varepsilon = 1/3\varepsilon_v$ and $\varepsilon = \varepsilon_v$ hold, where $\varepsilon_v = (W_0 - W)$ is the bulk strain caused by healing of the porosity. For uniaxial stress we have $\varepsilon \approx 5\varepsilon_v$ in the free state.

Thus, the degree of healing of porosity in copper under compression is determined by the value of the maximum shear stresses, the character of the distribution of stresses near the pore, and for shock-wave loading the duration of the compression pulse. The increase in efficiency of pore healing when dynamic methods of compression are used compared with static methods is probably due to the increase in the temperature of local heating and not to the change in stress state near the pore.

The work on shock-wave loading was supported by the Russian Fund for Fundamental Research under Grant No. 96-01-01207a.

[1] Ya. E. Geguzin and V. G. Kononenko, Fiz. Khim. Obrab. Mater., No. 2, 60 (1982).

[2] N. A. Zlatin, S. M. Monakov, G. S. Pugachev *et al.*, Zh. Tekh. Fiz. **45**, 681 (1975) [Sov. Phys. Tech. Phys. **20**, 428 (1975)].

[3] M. A. Meyers and L. E. Murr (Eds.), *Shock Waves and High-Strain-Rate Phenomena in Metals* (Plenum Press, New York, 1981; Metallurgiya, Moscow, 1984).

[4] E. L. Zil'berbrand, G. S. Pugachev, and A. B. Shinani, Pis'ma Zh. Tekh. Fiz. **20**(6), 45 (1994) [Sov. Tech. Phys. Lett. **20**, 236 (1994)].

[5] V. I. Betekhtin, A. B. Pakhomov, B. P. Peregud *et al.*, Zh. Tekh. Fiz. **59**(6), 136 (1989) [Sov. Phys. Tech. Phys. **34**, 668 (1989)].

[6] Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, Vols. 1 and 2 [Academic Press, New York, 1966/1967; Nauka, Moscow, 1966, 686 pp.].

Translated by Frank J. Crowne

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Nonholonomic constraint between the polarization state of light and the twisted angle of a single-mode optical fiber with linear birefringence

G. B. Malykin

Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia

Yu. I. Neĭmark

Scientific Research Institute for Applied Mathematics and Cybernetics, N. I. Lobachevski State University, Gorkiĭ

(Submitted October 1, 1997)
Zh. Tekh. Fiz. **68**, 128–129 (November 1998)

The relation between the polarization state of light propagating in a twisted single-mode optical fiber with linear birefringence and the twist angle of the fiber is discussed for the case where the fiber is stretched in a straight line. It is shown that this relation is of the nature of a nonholonomic constraint. © *1998 American Institute of Physics.* [S1063-7842(98)02711-1]

About 60 years ago, it was shown in a paper by Rytov[1] that when an optical beam propagates along a nonplanar path, the plane of its polarization rotates relative to the natural Darboux trihedron, which consists of the unit vectors tangent, normal, and binormal to the curved path of the beam. In this paper, Rytov showed that if the tangent vector returns to its original state at some point along the beam path, the plane of polarization of the light will in general differ from its original orientation, but that this phenomenon will not occur if the trajectory of the beam is a plane curve.

In Ref. 2, Berry showed that the Rytov effect can also occur in a single-mode optical fiber with a nonplanar configuration. In that paper,[2] he demonstrated that under these conditions the relation between the direction of the electric field vector and the spatial orientation of the Darboux trihedron, which is fixed in the optical fiber, is nonintegrable (nonholonomic).

The goal of this paper is to show that in an optical fiber with intrinsic linear birefringence, the presence of torsion (i.e., twisting of the fiber) gives rise to this same nonholonomic relation between the direction of the electric field vector in the fiber and the angle of twist of the fiber, even when the optical fiber is extended in a straight line. In other words, if only the state of polarization of the light at the input to the optical fiber and the orientations of the axes of its linear birefringence at the input α_1 and output ports α_2 are known, it is impossible in general to obtain a functional expression for the polarization state at the output of the optical fiber. Let us write the differential equations for the Jones vector

$$\mathbf{E}=\begin{vmatrix} E_x e^{i\psi_x} \\ E_y e^{i\psi_y} \end{vmatrix}$$

(see Ref. 3; E_x and E_y are amplitudes of the electric field components, and ψ_x and ψ_y are their phases) in the Cartesian system of coordinates that follows the twist of the optical fiber:

$$\frac{d\mathbf{E}}{dz}=N(z)\cdot\mathbf{E}(z). \tag{1}$$

Here z is the length measured from the beginning of the segment of optical fiber, and:

$$\mathbf{N}(z)=\begin{vmatrix} -i\beta_x & (1-g)\dfrac{d\alpha}{dz} \\ -(1-g)\dfrac{d\alpha}{dz} & -i\beta_y \end{vmatrix}$$

is the differential Jones matrix of an optical fiber[4] with intrinsic linear birefringence $\beta=\beta_x-\beta_y$, $\beta_{x,y}=(2\pi/\lambda)n_{x,y}$ (where λ is the wavelength of the light and n_x, n_y are the refractive indices for the slow and fast axes of the optical fiber; here we take n_x, $n_y=$ const) and a twist angle α; and $g=$ const is the photoelastic coefficient of the material from which the fiber was made.

The vector relation (1) can be written in scalar form using four real differential relations that express dE_x, dE_y, $d\psi_x$, and $d\psi_y$ in terms of dz and $d\alpha$. After a few simple transformations, we can write these relations as conditions under which the following four ordinary differential (Pfaffian) forms reduce to zero:

$$\begin{aligned} d\omega_{1,2}&=dE_{x,y}\mp(1-g)\cos(\psi_{x,x}-\psi_{y,y})d\alpha,\\ d\omega_{3,4}&=E_{x,y}d\psi_{x,y}+\beta_{x,y}E_{x,y}dz+(1-g)\\ &\quad\times E_{y,x}\sin(\psi_{x,x}-\psi_{y,y})d\alpha. \end{aligned} \tag{2}$$

For integrability of the differential equations $d\omega_1=d\omega_2=d\omega_3=d\omega_4=0$, i.e., in order that they be usable in principle to obtain finite relations that determine E_x, E_y, ψ_x and ψ_y as single-valued functions of z and α, it is necessary and sufficient that the bilinear forms $(\delta d-d\delta)\omega_s=0$ $(s=1,2,3,4)$ reduce identically to zero, together with $\delta\omega_s=0$ and $d\omega_s=0$.[5,6] The conditions $\delta\omega_s=0$ and $d\omega_s=0$ leave arbitrary the ‘‘variational’’ differentials δz, dz, $\delta\alpha$, and $d\alpha$. In this case the conditions of holonomic behavior, which reduce to the requirement that the bilinear form vanish identically for arbitrary δz, dz, $\delta\alpha$, and $d\alpha$, are not fulfilled, and consequently nonholonomic behavior occurs. This implies that when the variables z and α go from the same initial to the same final values along different paths, the resulting quantities E_x, E_y, ψ_x, and ψ_y will in general depend on the path, despite the identical initial values. An exception is the case $g = 1$. In real optical materials we have $0<g<1$. Thus, in quartz optical fibers $g = 0.08–0.065$.[7]

The nonholonomic nature of the relation between the polarization state of light propagating along an optical fiber and the twist in the fiber's birefringence axes can be illustrated with a simple example, in which the nonholonomic properties are a consequence of the noncommutative behavior of the Jones matrix for segments of optical fiber with different elliptical birefringences caused by twisting, for example, two segments of optical fiber with constant but different twisting of the axes.

Thus, the relation between changes in E_x, E_y, ψ_x, ψ_y and changes in z and α is nonholonomic in character, a phenomenon also encountered in theoretical mechanics,[5] the theory of electric motors and generators,[7] the theory of mechanical gyroscopes,[8] etc. Note that the propagation of light along a twisted optical fiber has a deep analogy with the time evolution of spherical coordinates of a sphere that is rolling without slipping along a plane while rotating around the binormal to the curve along which it is rolling on the plane.

Thus, for arbitrary relations between z and α, even if the state of optical polarization is known at the input to the optical fiber and the azimuth of the anisotropy axes is known at the input and output, it is impossible to compute the values of E_x, E_y, ψ_x and ψ_y that determine the state of polarization at the fiber output. If the function $\alpha(z)$ is given in explicit form, then, as is shown in Ref. 4, Eq. (1) can be reduced to a special Ricatti equation:

$$\frac{d\chi}{dz} = -n_{12}\chi^2 + (n_{22} - n_{11})\chi + n_{21},$$

where

$$\chi = \frac{|E_y|}{|E_x|} e^{i(\psi_y - \psi_x)},$$

and n_{ij} are elements of the differential Jones matrix.

This equation, as is well known, does not reduce to quadratures for an arbitrary form of $\alpha(z)$.

The axes of a linearly birefringent optical fiber become twisted both in the process of drawing it from the preform and in the process of installing it, e.g., into an optical communication line or the coil of an optical-fiber detector of some physical parameter. We note that a loop of optical fiber in a coil is already a nonplanar curve, within which changes in the polarization state of the light should occur connected not only with the twisting of the linear birefringence axes but also with the Rytov effect.[1]

In real cases the function $\alpha(z)$ in a fiber-optic communication line or coil cannot be found; only the values α_1 and α_2 at the beginning and end of the segment of optical fiber are known. As we showed above, in this case the state of polarization of the light cannot be found theoretically.

In conclusion, the authors are grateful to Ya. I. Khanin for discussing the results of this work, and to N. K. Vdovicheva and I. A. Shereshevskiĭ (Institute of Theoretical Mechanics, Russian Academy of Sciences) for useful consultations.

This work was supported in part by Grant No. 96-02-18568 from the Russian Fund for Fundamental Research.

[1] S. M. Rytov, Dokl. Akad. Nauk SSSR **18**, 263 (1938).
[2] M. V. Berry, Nature (London) **326**, 277 (1987).
[3] M. Monerie and L. Jeunhomme, Opt. Quantum Electron. **12**, 449 (1980).
[4] R. M. Azzam and N. M. Bashara, J. Opt. Soc. Am. **62**, 1252 (1972).
[5] Yu. I. Neĭmark and N. A. Fufaev, in *Dynamics of Nonholonomic Systems* [in Russian] (Nauka, Moscow, 1967), 519 pp.
[6] R. Ulrich and A. Simon, Appl. Opt. **18**, 2241 (1979).
[7] A. V. Gaponov, Dokl. Akad. Nauk SSSR **87**, 401 (1952).
[8] A. Yu. Ishlinskiĭ, *Mechanics of Gyroscopic Systems* (Israel Program for Scientific Translations, Jerusalem, 1965) [Russian original, Izd. AN SSSR, Moscow, 1963, 480 pp.].

Translated by Frank J. Crowne

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Simulation of arbitrary deformations of polycrystals by the method of cellular atomata

Ya. E. Beĭgel'zimer, A. V. Spuskanyuk, V. N. Varyukhin, and B. M. Éfros

A. A. Galkin Physico-Technical Institute at Donetsk, Ukrainian National Academy of Sciences, Donetsk, Ukraine

(Submitted October 13, 1997)
Zh. Tekh. Fiz. **68**, 130–132 (November 1998)

A computer model for the deformation of polycrystals is developed, based on the concept of cellular automata. The model is used to investigate a polycrystal by treating it as a multilevel hierarchical system, and to study the peculiarities in the behavior of this system as a function of the material structure, mechanisms of low-level deformation, history of loading and/or deformation at the top hierarchical level, the temperature, and other parameters. © *1998 American Institute of Physics.* [S1063-7842(98)02811-6]

1. In this paper we describe a class of model for deformable polycrystals based on cellular automata.[1,2] In essence, the concept of cellular automata reduces the region of study to a uniform grid, each cell of which contains several bits of data. As time progresses in discrete steps, the physical laws of behavior of the system are given by a unique set of rules (for example, a small handbook table), according to which every cell calculates its new state at each new time step based on the states of its nearest neighbors. The authors of Ref. 3 proposed to use the method of moving cellular automata to simulate the deformation of solids.

The original solution treated by these authors was a set of cellular automata with self-similar structure, which made it possible to immediately build into the model the fractal structure of real materials (it is known, for example, that the structure of real steels is fractal[4]). Moreover, such self-similar structures are adequately realized by the methods of object-oriented programming, which makes it possible to use the most recent developments in programming language for numerical simulations, ensuring a high-speed and rational allocation of computer memory, clarity of programming, and the possibility of easy modification to accommodate new mechanisms of deformation and material structure.

2. We will model the structure of a real polycrystal[4] by a three-dimensional cellular structure. The cells can be simple or compound. Simple cells have no internal structure, whereas compound cells consist of simple and/or compound subcells. As a component, a compound cell can contain cells similar to itself, which allows us to model the fractal structure.

In this paper we use compound cells having the structure of a cubic lattice and consisting of 27 ($3\times3\times3$) cells of smaller sizes (Fig. 1). In general, other spatial structures are possible as well as other numbers of components.

Let us introduce the environment of a cell into the discussion, by which we mean its set of nearest neighbors, and assign to each cell three coordinates (m,n,k) that determine its position (where m,n,k are integers from 1 to 3). In Fig. 3 the central cell has coordinates $(2,2,2)$, and its neighborhood consists of all the remaining 26 cells. In defining the neighborhood of the boundary cells we specify that the system of cells in Fig. 1 is surrounded on all sides by cells that are similar to it. This is the case of so-called periodic boundary conditions, according to which the environment of each of the 27 cells derives from the remaining 26 cells.

3. A loaded polycrystal is characterized by nonuniform stress–strain states. In order to describe them we introduce stress σ^n and plastic strain tensors e_p^n into the discussion for each cell of the structure described above. In the notation we will use here, the label n indicates the level to which the given cell belongs. In this case we assume that the largest cell, which models the representative volume of the macroscopic level, belongs to level 1. Its 27 component cells belong to level 2, the components of each of these 27 cells belongs to level 3, etc. It is obvious that at level n there are 27^{n-1} cells.

Because plastic strain of a cell of the nth level is caused by plastic strains of its component cells at the $(n+1)$-th level, we assume that

$$e_p^n=\langle e_p^{n+1}\rangle, \tag{1}$$

where the angle brackets indicate averaging over the volume of the nth level cell.

In what follows, we will use the notation of direct tensor calculus.

Let us assume that the stress and plastic-strain tensors of a cells of level n and its constituent 27 cells of level $n+1$ are related by the Kröner equation

$$\sigma^{n+1}-\sigma^n=M(e_p^n-e_p^{n+1}), \tag{2}$$

where M is in general a fourth-rank tensor.

In this paper we assume that M is a scalar quantity, and in what follows we will refer to it as the accommodation parameter, the meaning of which will become clear as the results of the computer experiment are discussed.

The previous expression shows that the difference in plastic deformation of each cell of level $n+1$ from the average value of this deformation within the corresponding cell of the nth level gives rise to internal microstresses that tend to smooth out this deformation. In this case there is a redistribution of the stresses within the nth-level cell, and a nonuniform stress–strain state arises within the compound cells.

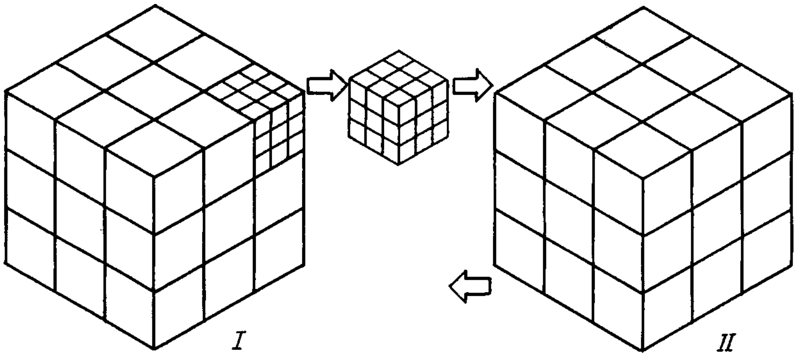

FIG. 1. Structure of the cellular automata. The first-level cell (*I*) equals 27 smaller cells of the second level (*II*).

Relation (1) allows us to find the plastic deformation of compound cells in terms of plastic deformation of their components. The plastic deformation of simple cells is determined by the deformation mechanisms acting within them.

In particular, for plastic deformation that comes from slipping of dislocations, the value of the rate of plastic deformation is calculated by summing up the components caused by all the slip systems that act within the cell. In this case, for small elastoplastic deformations the quantity $\dot{e}_p$ is calculated from the expression

$$\dot{e}_p=\frac{1}{2}\sum_{\alpha}\ \dot{\gamma}^{\alpha}(s^{\alpha}m^{\alpha}+m^{\alpha}s^{\alpha}), \tag{3}$$

where m^{α} and s^{α} are respectively vectors normal to the slip plane and the direction of slip in the system α, while $\dot{\gamma}^{\alpha}$ is the rate of shear strain in system α.

The value of $\dot{\gamma}^{\alpha}$ is determined by the tangential stresses τ^{α} acting on the system α, and the value of τ^{α} is calculated as usual from the stress tensor of the corresponding cells:

$$\tau^{\alpha}=m^{\alpha}s^{\alpha}:\sigma^{\alpha}. \tag{4}$$

The relations that connect $\dot{\gamma}^{\alpha}$ and τ^{α} for various mechanisms that control the motion of dislocations are described in a whole series of publications on the physics of plastic deformation. In particular, according to Eshby,[5]

$$\dot{\gamma}^{\alpha}=\dot{\gamma}_0^{\alpha}\exp\left(-\frac{\Delta F}{kT}\left(1-\left(\frac{\tau^{\alpha}}{\tau_c^{\alpha}}\right)^{p}\right)^{q}\right), \tag{5}$$

where ΔF is the activation energy required to overcome obstacles in the absence of external stresses; τ_c^{α} is the critical tangential stress for system α; p and q are parameters that depend on the mechanism that controls the motion of dislocations $(0\leqslant p\leqslant 1, 1\leqslant q\leqslant 2)$; k is Boltzmann's constant; T is the temperature; and $\dot{\gamma}_0^{\alpha}$ is a certain parameter that characterizes the system α.

In accordance with the concept of cellular automata, we will investigate the generation of the stress–strain state in discrete time t_m with discretization step Δt (where $t_m=m\Delta t$ and m is an integer). At the top level we prespecify the time dependence of the stresses acting: $\sigma^1=\sigma^1(t_m)$. At the initial time $(m=0)$ we will assume that the plastic deformations e_p^{α} equal zero at all levels.

Relation (2) we write in a somewhat modified form

$$\sigma^{n+1}(t_m)-\sigma^{n}(t_m)=M(e_p^{n}(t_{m-1})-e_p^{n+1}(t_{m-1})), \tag{6}$$

which allows us to calculate the cellular stresses at time t_m based on the plastic deformations at the previous time t_{m-1}.

Relations (1)–(6) make it possible to determine the stress–strain state of a polycrystal according to a given program of loading at the top level.

If the history of straining the material at the top level is given, i.e., the total strain tensor $e^1=e^1(t_m)$ is specified, then the quantity $\sigma^1(t_m)$ is determined, according to Hooke's law

$$\sigma^1(t_m)=E:e_e^1(t_m), \tag{7}$$

as a function of the elastic strain $e_e^1(t_m)$, where E is the tensor of elastic constants.

The value of the elastic strain we find from the expression

$$e_e^1(t_m)=e^1(t_m)-e_p^1(t_{m-1}). \tag{8}$$

Relations (1)–(8) make it possible to determine the stress–strain state of a polycrystal according to a prespecified strain history at the top level.

4. As an example we present the results of computer simulation experiments of compound strain in α-iron polycrystals. The parameters of Eq. (5) were taken from the book by Ashby:[5] $\Delta F=0.5\mu_0 b^3$, $b=2.48\times10^{-10}$ m is the modulus of the Burgers vector, $\mu_0=6.4\times10^{10}$ Pa is the shear modulus, and $\tau_c=1.7\times10^{-3}\mu_0$.

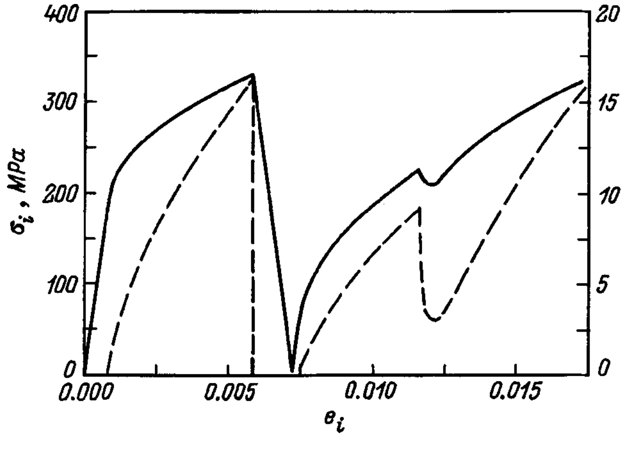

FIG. 2. Hardness curves and fraction of plastically deformed elements N for compound loading of an α-Fe polycrystal: solid curve — σ, dashed curve — N.

TABLE I. Program for compound straining.

Step No.	e_{xx}	e_{yy}	e_{zz}	e_{xy}	e_{yz}	e_{zx}
1	0	0	0	0	0	0
2	0.005	-0.005	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0.005	0	0

Figure 2 shows the computed hardening curve for the deformation history of the polycrystal listed in Table I.

In Table I e_{ij} are the components of the total strain tensor of a macroelement of the material (a cell at the top level). We call attention to the following. During unloading and loading in the opposite direction (step 3) the Baushinger effect appears. In the course of the shear deformation, at step 4 a time-dependent unloading of the material takes place, after which the hardening curve approaches the continuation of the curve of the previous step. In this we see manifested the well-known phenomenon of the existence of a common hardening curve.

Figure 3 shows the results of calculations for strain in the direction of the second step of Table I for various values of the parameter M in Eq. (2). According to this equation, the physical meaning of M is to incorporate into the calculation a capability of the structural elements to accommodate each other. The larger M is, the less able these structures are to accommodate, i.e., the stronger the reaction of the environment to a "stranger." It follows from Fig. 3 that the larger M is, the higher the hardness curve will go, other conditions being equal.

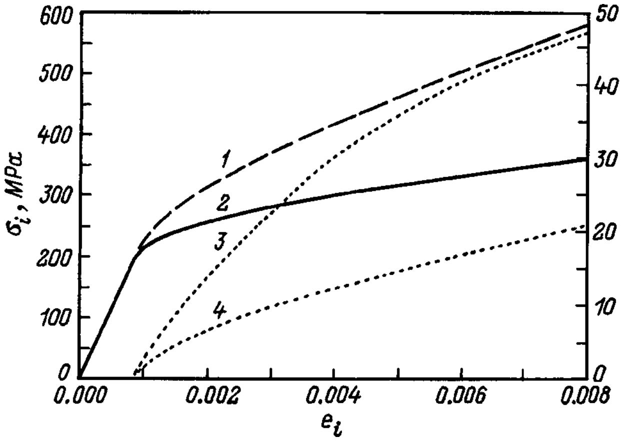

FIG. 3. Effect of the accommodation coefficient M on the hardening curve and the number of plastically deformed elements N (*1,2* — σ; *3,4* — N); $M=1000$ (*2,4*) and 5000 MPa (*1,3*).

[1] T. Toffoli and N. Margolus, *Cellular Automata Machines: A New Environment for Modeling* (MIT Press, Cambridge, Mass., 1987; Mir, Moscow, 1991, 280 pp.).

[2] K. Preston and M. Duff, *Modern Cellular Automata. Theory and Applications* (Plenum Press, New York, 1994), 327 pp.

[3] S. G. Psakh'e, A. Yu. Smolin, S. Yu. Korostelyov *et al.*, Pis'ma Zh. Tekh. Fiz. **21**(20), 72 (1995) [Tech. Phys. Lett. **21**, 849 (1995)].

[4] E. Hornbogen, Prakt. Metallogr. **23**, 258 (1986).

[5] H. J. Frost and M. F. Ashby, *Deformation-Mechanism Maps: The Plasticity and Creep of Metals and Ceramics* (Pergamon Press, Oxford, 1982), 425 pp.

Translated by Frank J. Crowne

Influence of betatron oscillations on magnetic bremsstrahlung

O. E. Shishanin

Moscow State Industrial University, 109280 Moscow, Russia

(Submitted November 4, 1997)
Zh. Tekh. Fiz. **68**, 133–134 (November 1998)

A description of the motion of an electron in an accelerator is discussed in relation to the problem of synchrotron radiation. Spectral–angular distributions of the radiation intensity are obtained that depend strongly on vertical oscillations of the electron. © *1998 American Institute of Physics.* [S1063-7842(98)02911-0]

As the authors of Ref. 1 emphasized, it is only the integrated intensity and the spectral distribution of the synchrotron radiation that remain almost unchanged in a storage ring. The first discussions of the effect of betatron oscillations on the angular properties of synchrotron radiation in an axially symmetric magnetic field appeared in Ref. 2. Subsequently this problem was investigated for a strongly focusing system, and also for weak focusing.[3–5]

It is well known[6] that vertical betatron oscillations of charged particles in accelerators can be described by the function

$$z=\sqrt{\frac{\beta_z A}{\pi}}\cos\left(\int\frac{ds}{\beta_z}+\delta_0\right), \tag{1}$$

where A is the emittance, β_z is the betatron function, which depends on the length s of the orbit, and δ_0 is the initial phase.

Let $\varphi=s/R_0$ be the generalized azimuth, where, in the case of magnet systems with rectilinear gaps, R_0 will be the average radius (the length of the orbit divided by 2π). Then the corresponding velocity component from Eq. (1) is given by

$$v_z=c\sqrt{\frac{A}{\pi\beta_z}}\sqrt{1+\left(\frac{1}{2}\frac{d\beta_z}{ds}\right)^2}\cos\left(\int\frac{ds}{\beta_z}+\delta_1+\delta_0\right), \tag{2}$$

where $\sin\delta_1=1/\sqrt{1+(d\beta_z/ds)^2/4}$.

Since radial oscillations have only a small effect on the angular distribution of the radiation, we can assume that a particle moves along a circle of radius R_0 and that the magnetic field in the bending magnets may be averaged over a period. Thus, the position of a particle will be determined by the radius vector

$$\mathbf{r}=\{(R_0+\rho)\cos\varphi,\ (R_0+\rho)\sin\varphi,\ z\}.$$

Expanding the transverse components of the magnetic field in a Fourier series, we find for the angular velocity the following expression

$$\dot{\varphi}=\frac{\omega_0}{1+k}\left[1-\frac{\rho}{R_0}+\frac{3}{2}\frac{\rho^2}{R_0^2}+\frac{1}{R^2}\int(z\dot{z}-\rho\dot{\rho})n(\varphi)dt\right], \tag{3}$$

where $\omega_0=eH/mc$, k is the ratio of the mean free path of a particle to the lengths of the bending magnets, $R_0=(1+k)R$, and the periodic function $n(\varphi)$ has a specific form for a given magnetic structure.

In studying the properties of synchronous radiation we will follow the operational method of Ref. 2. Let the radiation vector $\mathbf{k}=\omega\mathbf{n}/c$, where $\mathbf{n}=\{0,\sin\theta,\cos\theta\}$, and θ is the spherical angle. The linear polarization vector we define as

$$\mathbf{e}_\sigma=\{1,0,0\},\quad \mathbf{e}_\pi=\{0,\cos\theta,-\sin\theta\}.$$

Then the components of the radiated intensity in the first-quantization approximation can be written in the form

$$\frac{dW_\sigma}{d^3k}=W_1\left|\int dt v_x\exp\left(i\frac{\nu'}{\nu}(\omega t-\mathbf{k}\cdot\mathbf{r})\right)\right|^2,$$

$$\frac{dW_\pi}{d^3k}=W_1\left|\int dt(v_y\cos\theta-v_z\sin\theta)\exp\left(i\frac{\nu'}{\nu}(\omega t-\mathbf{k}\cdot\mathbf{r})\right)\right|^2, \tag{4}$$

where

$$W_1=\frac{ce^2}{(2\pi)^3R_0}\frac{\nu'}{\nu},\quad \omega=\nu\frac{\omega_0}{1+k},\quad \nu'=\nu\left(1+\frac{h\omega}{E}\right).$$

The angle φ can be defined in reference to any point on the orbit. Then the small parameters will be $\varphi\sim m_0c^2/E$, $\tau=N\varphi$ (where N is the number of magnetic periods), and also $\cos\theta(\theta\sim\pi/2)$, ρ/R_0, z/R_0. Taking this into account in

$$\omega t-\mathbf{k}\cdot\mathbf{r}=\nu\frac{\omega_0}{1+k}\left[t-\frac{1}{c}(R_0+\rho)\sin\varphi\sin\theta-\frac{z}{c}\cos\theta\right]$$

we set $\sin\varphi\approx\varphi-\varphi^3/6$, $\sin\theta\sim1$.

In order to find φ in Eq. (3) we must carry out the integration. In particular, to zero approximation we have $\varphi=\omega_0t/(1+k)$. From the equation $v^2=\dot{\rho}^2+\dot{z}^2+r^2\dot{\varphi}^2$ we find the value of $R\omega_0/c$, while the velocity components are determined from $v_x\approx\dot{\rho}-c\varphi$, $v_y\approx c\beta$, $v_z=\dot{z}$.

Let us make the substitution $\xi=-v_x/c=\varphi-\dot{\rho}/c$, and then carry out an expansion of the type

$$z=z\Big|_{\tau=0}+\frac{dz}{d\tau}\Big|_{\tau=0}\cdot\tau+\ldots\approx z_0+R_0\frac{v_z}{c}\varphi.$$

To sum up, retaining terms to third order of smallness, we obtain for $\omega t-\mathbf{k}\cdot\mathbf{r}$ the expression

$$\nu\xi\left[1-\beta\sin\theta+\frac{\xi^2}{6}-\frac{v_z}{c}\cos\theta+\frac{1}{2}\left(\frac{v_z}{c}\right)^2\right]+\text{const.}$$

Then in the ultrarelativistic limit we have $1-\beta\sin\theta\approx\varepsilon/2$, where $\varepsilon=1-\beta^2\sin^2\theta$.

Performing the integration in Eq. (4) with respect to the new variable ξ, and also averaging over initial phases, we obtain the spectral–angular distribution in the following form:

$$\frac{dW_\sigma(\nu)}{d\Omega}=W_2\int_0^{2\pi}d\delta\varepsilon_1^2K_{2/3}^2\left(\frac{\nu'}{3}\varepsilon_1^{3/2}\right),$$

$$\frac{dW_\pi(\nu)}{d\Omega}=W_2\int_0^{2\pi}d\delta\varepsilon_1\varepsilon_2K_{1/3}^2\left(\frac{\nu'}{3}\varepsilon_1^{3/2}\right),\tag{5}$$

where

$$W_2=\frac{ce^2\nu\nu'}{12\pi^4R_0^2},\quad \varepsilon_1=1-\beta^2+\varepsilon_2,$$

$$\varepsilon_2=(\cos\theta-\alpha\cos\delta)^2,$$

$$\alpha=\sqrt{\frac{A}{\pi}}\left[\frac{1}{\sqrt{\beta_z}}\sqrt{1+\left(\frac{1}{2}\frac{d\beta_z}{ds}\right)^2}\right]_{\tau=0}.$$

If we introduce the angle ψ measured from the plane of the orbit, then $\cos\theta$ in Eq. (5) can be replaced by ψ. If we sum over the spectrum in (5), then the angular distributions can be expressed in terms of elliptic integrals[4] and will also depend on the vertical motion of the electron.

Expression (5) is also convenient for storage rings. In using it for practical applications, we must keep in mind that usually a plot of the betatron function β_z is known for all the orbits. Then its value must be taken for that point from which the radiation is emitted, and its derivative may be approximately replaced by the ratio $\Delta\beta_z/\Delta s$. For bending magnets the plot of the β_z function is approximately a straight line (see, e.g., Fig. 3 in Ref. 7); therefore, the slope of this portion of the line can also be used.

The algorithm for computing the integrals (5) is described in particular in Ref. 3. In the cases of most interest (small oscillation amplitudes or small beam sizes) we can expand in the parameter $q^2=\alpha^2/2\varepsilon$, where $\varepsilon\approx1-\beta^2\sin^2\theta=(1+\gamma^2\psi^2)/\gamma^2$.

Then in the classical limit the right-hand sides of (5) can be replaced respectively by

$$W_3[K_{2/3}^2+q^2\varepsilon^2\psi^2\nu^2(K_{1/3}^2+K_{2/3}^2)-q^2\varepsilon\sqrt{\varepsilon}\,\nu(1+2\psi^2/\varepsilon)K_{1/3}K_{2/3}],$$

$$W_3[(\psi^2/\varepsilon+q^2)K_{1/3}^2+q^2\varepsilon\psi^4\nu^2(K_{1/3}^2+K_{2/3}^2)-5q^2\sqrt{\varepsilon}\,\psi^2\nu K_{1/3}K_{2/3}],\tag{6}$$

where

$$W_3=\frac{ce^2\nu^2\varepsilon^2}{6\pi^3R_0^2},\quad K_i=K_i(\nu\varepsilon^{3/2}/3).$$

The well-known expressions for a uniform magnetic field can be recovered here by setting $q^2=0$ (no vertical oscillations) and replacing R by R_0.

If we keep $\cos\theta$ instead of ψ in Eq. (6) and carry out the integration over the spherical angle, then the spectral expressions obtained coincide with the expressions for a uniform magnetic field, in agreement with Ref. 1.

The plots constructed according to Eq. (5) show that the maxima of both components will be lower than the curves for a uniform magnetic field (for the same energy E and radius R). The radiation in the plane of the orbit will no longer be completely linearly polarized. For the π component a maximum appears at $\theta=\pi/2$ instead of the minimum in the high-amplitude case; it is also possible for the σ component to exhibit a slight minimum in the orbital plane at extremely high amplitudes, and the π component can develop additional symmetric local depressions. Thus, along with theoretical studies of this effect, it is important to carry out more-precise experiments on modern accelerators. Since the radiation is primarily incoherent, the results obtained will give a good description of the properties of the radiation generated by beams of electrons as well.

[1] G. N. Kulipanov and A. N. Skrinskiĭ, Usp. Fiz. Nauk **122**, 369 (1977) [Sov. Phys. Usp. **20**, 559 (1977)].

[2] V. Ch. Zhukovskiĭ and O. E. Shishanin, Zh. Éksp. Teor. Fiz. **61**, 1371 (1971) [Sov. Phys. JETP **34**, 729 (1971)].

[3] O. E. Shishanin, Zh. Éksp. Teor. Fiz. **103**, 1117 (1993) [Sov. Phys. JETP **76**, 547 (1993)].

[4] O. E. Shishanin, JETP Lett. **57**, 785 (1993).

[5] O. E. Shishanin, Teor. Mat. Fiz. **106**, 285 (1996).

[6] N. S. Dikanskiĭ and D. V. Pestrikov, *Physics of High-Current Beams and Storage Rings* [in Russian] (Nauka, Novosibirsk, 1989), 333 pp.

[7] A. G. Valentinov, P. D. Voblyĭ *et al.*, IYaF AN SSSR Preprint No. 89-174 [in Russian], Institute of Nuclear Physics, Siberian Branch of the Academy of Sciences of the USSR, Novosibirsk (1989), 30 pp.

Translated by Frank. J. Crowne

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Thermally stimulated flow of low-melting glasses during planarization of the surface profiles of microelectronic structures

S. P. Zhvavyĭ, G. D. Ivlev, V. A. Pilipenko, and V. N. Ponomar'

Electronics Institute, Academy of Sciences of Belarus, 220090 Minsk, Belarus

(Submitted November 4, 1997)
Zh. Tekh. Fiz. **68**, 135–137 (November 1998)

The problem of characterizing the flow of phosphorosilicate and borophosphorosilicate glasses during thermally stimulated planarization of the surface profile of microelectronic structures is solved in a model based on Newtonian viscous flow driven by surface tension forces. Expressions are obtained for the melting angle as a function of the temperature–time regimes of the heat treatment and the physical parameters of the glasses. © *1998 American Institute of Physics.* [S1063-7842(98)03011-6]

Phosphorosilicate and borophosphorosilicate glasses are widely used in integrated-circuit technology to make the interlevel insulating layers of microelectronic structures. It is possible to smooth out sharp projections at the edges of contact windows and to planarize the surface profile of these structures by melting the glass. Although the melting of phosphorosilicate and borophosphorosilicate glasses has been studied in a number of papers,[1–5] the data obtained leave a number of questions unanswered, notably how the melting angle of a step made of phosphorosilicate (or borophosphorosilicate) glass depends functionally on the heat-treatment regime and physical parameters of the glasses. Another question without a definitive answer is, why is it that melting induced by second-long optical heating is so much more efficient than long periods of isothermal annealing in an oven?

In order to describe the melting of a glass step (Fig. 1) we start with a model that assumes Newtonian viscous flow driven by surface tension forces. In this case

$$\mu \frac{\partial v_x}{\partial y} = \frac{\sigma}{r}, \tag{1}$$

where μ is the viscosity, v_x is the flow velocity along the x axis, σ is the surface tension, and r is the radius of curvature.

The flow of the glass is caused by the tangential stress $P_{xx} = -\sigma/r$. Curve AB (Fig. 1) can be treated[4] as a half period of a sine function with period $2x$ and amplitude $h/2$. P_{xx} is a maximum when

$$r = \frac{x^2}{2\pi^2 h} = \frac{h}{2\pi^2}\cot^2\alpha. \tag{2}$$

The horizontal flow velocity field of a fluid with a "wedge" profile can be written in the form[6]

$$v_x(y) = \frac{3(2hy - y^2)}{2h^2} u, \tag{3}$$

where u is the average flow velocity.

Substituting Eq. (3) into Eq. (1), we find that the average translation velocity of point B (Fig. 1) is

$$u = \frac{h}{3\mu}\,\frac{\sigma}{r}. \tag{4}$$

On the other hand, we can also write the following expression:

$$u = \frac{dz}{dt} = -\frac{h}{\sin^2\alpha}\,\frac{d\alpha}{dt}. \tag{5}$$

Substituting Eqs. (2) and (5) into Eq. (4) we obtain the equation

$$-\frac{\cos^2\alpha}{\sin^4\alpha}\,\frac{d\alpha}{dt} = \frac{2\pi^2}{3}\,\frac{\sigma}{\mu h}, \tag{6}$$

whose solution is

$$\cot^3\alpha - \cot^3\alpha_0 = \frac{2\pi^2}{h}\int_{t_1}^{t_2} \frac{\sigma}{\mu}\,dt. \tag{7}$$

It follows from Eq. (7) that the dependence of the melting angle on temperature is determined by the functions $\sigma(T)$ and $\mu(T) = \mu_0 \exp(E/RT)$, where E is the activation energy for viscous flow and R is the universal gas constant. Since the temperature dependence of $\sigma(T)$ is quite weak ($d\sigma/dT \approx 0.12$ erg/cm$^2 \cdot$K; see Ref. 7), we can assume that $\sigma =$ const. For isothermal annealing we obtain from Eq. (7)

$$\cot^3\alpha - \cot^3\alpha_0 = 2\pi^2 \frac{\sigma}{\mu_0 h} t \exp(-E/RT). \tag{8}$$

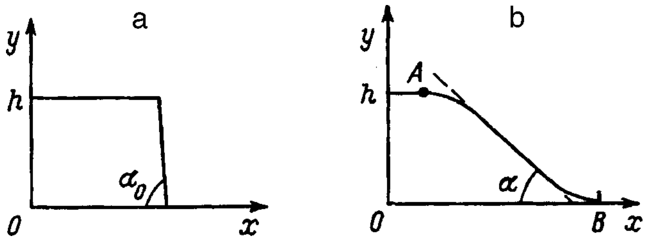

FIG. 1. Profile of a phosphorosilicate glass step before (a) and after (b) melting.

It follows from Eq. (8) that cot $\alpha \sim t^{1/m}$, where $m = 3$. In an analogous empirical relation derived in Refs. 2 and 3 and based on experimental data for borophosphorosilicate glasses, the quantity m was 2.4 ± 0.8, which coincides with the value of 3 we have obtained within the limits of possible error.

During second-long heat treatment, the substrate temperature at the end of the optical pulse increases to a maximum T^*, and then drops rapidly.[8] The increase in T and subsequent decrease are approximated by writing $T(t) = T^* \sqrt{t/\tau}$ and $T(t) = T^* \sqrt{\tau/t}$ respectively, where τ is the heating time, i.e., the duration of the optical pulse. From these expressions we derive the limits of integration in Eq. (7): $t_1 = (T_g/T^*)^2\tau$ and $t_2 = (T^*/T_g)^2\tau$, where T_g is the glass-point (softening) temperature, at which $\mu = 10^{13}$ Pa.

In general[9–11] one has $E(T) = E_0 - TS$, where E_0 and $S = -\partial E/\partial T$ are the enthalpy and entropy of activation of the viscous flow, respectively. If the function $E(T)$ is approximately linear over the temperature range from T_g to T^*, i.e., if we assume that E_0 and S are average parameters, then we can find an analytic solution to Eq. (7) that describes the change in the angle α for pulsed heat treatment of the glasses:

$$\cot^3\alpha - \cot^3\alpha_0 = 2\pi^2 \frac{\sigma}{\mu_0 h}\tau \cdot e^{S/R}\left\{ e^{-a}\left[1 - a + \frac{2}{a}\left(1 + \frac{1}{a}\right)\right] - a^2 \cdot \mathrm{Ei}(-a) - e^{-a_g}\left[\left(\frac{T_g}{T^*}\right)^2 (1 - a_g) + \frac{2}{a}\left(\frac{T^*}{T_g} + \frac{1}{a}\right)\right] + a^2 \cdot \mathrm{Ei}(-a_g)\right\}, \qquad (9)$$

where $a_g = E_0/RT_g$, $a = E_0/RT^*$, and $-\mathrm{Ei}(-x) = \int_1^\infty e^{-xy} y^{-1} dy$ is the exponential integral function.

Since $E/RT \sim 30 \gg 1$, expanding $-\mathrm{Ei}(-x)$ in a series and keeping terms proportional to RT/E_0 we obtain from Eq. (9)

$$\cot^3\alpha - \cot^3\alpha_0 = 8\pi^2 \frac{\sigma}{\mu_0 h} \frac{RT^*}{E_0} \tau \exp\left(-\frac{E_0 - T^* S}{RT^*}\right). \qquad (10)$$

In carrying out calculations based on Eqs. (8) and (10) (Fig. 2) for the specific experimental data from Ref. 1, we use values $\mu_0 = 10^{-3}$ Pa and $\sigma = 400$ erg/cm^2 (Ref. 7). The calculated functions $\alpha(T)$ for isothermal melting of borophosphorosilicate glass ($t = 60$ min) are in rather good agreement with experiment if we pick $E = 300$ (curve *1*) and 280 kJ/mol (curve *2*), except for the values of α at $T = 900$ °C. For pulsed heat treatment of the glasses ($\tau = 10$ s), the temperature dependences of the melting angle were obtained by choosing $E_0 = 325$ kJ/mol and $S = 10$ J/mol·K (curve *3*) and $E_0 = 285$ kJ/mol and $S = 20$ J/mol·K (curve *4*). The experimental values of $\alpha(T^*)$ are extremely close to the calculated values or coincide with them. It is worth noting that the average values E, E_0, and S we have found agree with handbook data for the various glass compositions.[7] Comparing these results with measured data confirms the correctness of our analytical expressions for α.

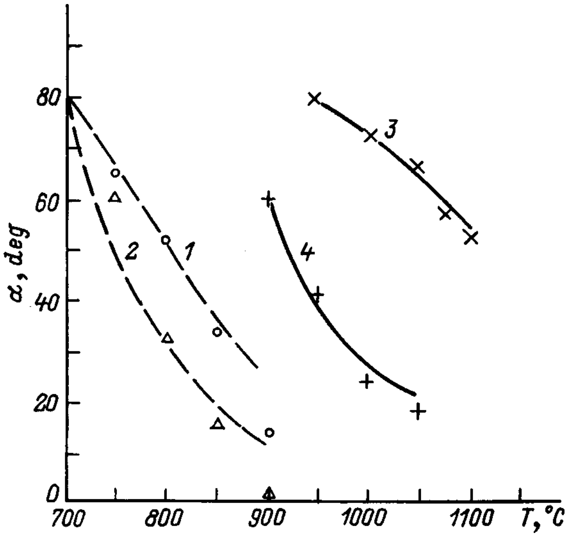

FIG. 2. Temperature dependence of the slope angle of a borophosphorosilicate glass step for the following concentrations of doping impurities: *1* — 4.6% B and 3.1% P, *2* — 5.6% B and 3.1% P, *3* — 2.5% B and 3.6% P, *4* — 6.3% B and 2.9% P. The dashed curves correspond to isothermal annealing, the solid curves to pulsed heating.

Thus, our investigations show that the expressions we have obtained for the temperature–time dependences of the melting angle of steps made of phosphorosilicate (borophosphorosilicate) glass on the basis of a model of Newtonian viscous flow under the action of surface tension forces give a rather good description of the melting process. By including the temperature dependence of the activation energy for viscous flow we were able to obtain expressions for the melting angle under pulsed heating and to show that the efficiency of this process is a consequence of the dynamic decrease in viscosity due to the entropy factor.

[1] F. S. Becker, D. Pawlik, H. Schäfer, and G. Staudigl, J. Vac. Sci. Technol. B **4**, 732 (1986).

[2] J. S. Mercier, Solid State Technol. **30**(7), 90 (1987).

[3] V. Yu. Vasil'ev and T. G. Dukhanov, Élektron. Tekh. Ser. 3. Mikroélektronika, No. 4, 38 (1991).

[4] L. K. White, N. A. Miszkowski, W. A. Kurylo, and J. M. Shaw, J. Electrochem. Soc. **139**, 822 (1992).

[5] G. Thallikar, H. Liao, and T. S. Cale, J. Vac. Sci. Technol. B **13**, 1875 (1995).

[6] P. G. de Gennes, Rev. Mod. Phys. **57**, 827 (1985); Usp. Fiz. Nauk **151**, 619 (1987).

[7] O. V. Mazurin, M. V. Strel'tsyna, and T. P. Shvaĭko-Shvaĭkovskaya, *Properties of Glasses and Glass-Forming Melts (a Handbook)* [in Russian], Vol. 1 (Nauka, Leningrad, 1973), 444 pp.

[8] V. E. Borisenko, S. N. Kornilov, V. A. Labunov, and Yu. V. Kucherenko, Zarubezh. Élektron. Tekh. **6**, 45 (1985).

[9] S. V. Nemilov, Fiz. Khim. Stekla **4**(2), 129 (1978).

[10] S. V. Nemilov and N. G. Gutkina, Fiz. Khim. Stekla **6**, 535 (1980).

[11] R. L. Myuller, Zh. Prikl. Khim. **28**, 1077 (1955).

Translated by Frank J. Crowne

TECHNICAL PHYSICS VOLUME 43, NUMBER 11 NOVEMBER 1998

Identification of the bow stabilization mechanism by numerical simulation of the laminar asymmetric flow of a viscous incompressible fluid past a cylinder with a projecting disk

S. V. Guvernyuk, S. A. Isaev, and A. G. Sudakov

Civil Aviation Academy, 196210 St. Petersburg, Russia

(Submitted February 9, 1998)
Zh. Tekh. Fiz. **68**, 138–142 (November 1998)

The mechanism responsible for the bow stabilization that arises as a result of the deformation of the spatial vortex structures that are organized in the flow of a fluid past a cylinder with a projecting thin disk is analyzed by solving the Navier–Stokes equations using a factorized finite-volume method and a simplified approach based on analytical specification of the metric coefficients. © *1998 American Institute of Physics.* [S1063-7842(98)03111-0]

The idea of organizing the flow around bodies by the deliberate placement of projecting thin plates and disks that generate large-scale vortex structures, which was proposed by Belov[1] more than 20 years ago, has turned out to be an exceptionally fruitful and successful approach for reducing the aerodynamic drag of bluff bodies and significantly increasing their static stability (the latter effect has been given the name ‘‘bow stabilization’’). By mounting a projecting coaxial disk on the body, one can reduce the wave drag of a blunt body in a supersonic flow by a factor of 2 to 4 (Ref. 2) or lower the profile drag of a cylinder in an axisymmetric turbulent flow of incompressible fluid by almost two orders of magnitude.[3,4]

The problem of reducing the drag on bodies in an organized separated flow, which is the first part of the complex problem of choosing an aerodynamic figure for bodies of unconventional shape, has received detailed and successful solution through numerical and physical simulations. A far from complete list of useful results deriving from systematic investigations of this class of problems would include: the existence of optimal ways of mounting a disk on a high-aspect cylinder so as to minimize the frontal and profile drag coefficients in the laminar-flow regime,[5] estimation of how strongly the degree of turbulence of an oncoming flow influences the drag of a body with a forward separation zone,[6] extension of the concept of organized flow separation to the flow in the near wake of a body, and the search for optimal ways of mounting protruding disks on bluff cylinders to give drag coefficients C_x comparble to those for bodies of streamlined shape.[7]

The first study of the problem of static stability of bodies with organized large-scale vortex structures was undertaken to address the problem of flow past loads suspended from helicopters.[8,9] The engineering solution to this practical, real-world problem was to choose suitable dimensions and positions for a shield placed in front of the end of a parallelepiped, which simulates a load such as a girder or container. However, it proved possible to ensure flight stability of bodies with this arrangement only by using tail stabilizers. As a result of a combined approach based on calculations, wind-tunnel experiments, and flight trials, the proposed solutions were validated; however, this approach did not adequately elucidate the nature of bow stabilization, nor the hydrodynamic aspects of how it acts as a control mechanism. To a certain extent this was due to the extreme computational difficulty of dealing with problems of spatial flow past bodies with unconventional shapes and organized separation zones. The glaring lack of information regarding bow stabilization and the need for more qualitative and exact resolution of this problem dictates a return to the classical problem of flow of a stream of viscous incompressible fluid past a semi-infinite cylinder with a disk at nonzero angle of attack, with enhancement of the detail of the solution within the leading separation zone.

In this numerical investigation a central place is occupied by a detailed analysis of the control mechanism for bow stabilization and identification of spatial vortex structures by using the Tecplot graphics package to interpret the results of the numerical calculations, thereby providing computer visualization of flows spreading along the control surfaces, and observation of particles following the tracks of tracer liquid particles. We will consider a structure made up of a semi-infinite cylinder and a protruding disk (diameter $d=0.75$, gap width $l=0.375$) that is optimum with respect to profile drag (at high Reynolds numbers[3,4]) in stable laminar-flow regimes with Re = 700 and 1000, with the formation of a leading separation zone at angles of attack of 0 and 5°. The diameter of the connecting column was 0.18. We took the diameter of the cylinder and the velocity of the oncoming flow as the characteristic scales.

A previous approach[9] to the numerical simulation of separated flow past a load employed *H*-shaped and in genreral nonorthogonal meshes matched to the surface of the body, with the mesh points arranged in planes perpendicular and parallel to the oncoming flow and having a higher density near the wall and in zones where shear layers develop. In contrast to that approach, in our investigation we make use of a cylindrical mesh that is matched with the washed surface. This allows us to increase considerably (by a factor of 5 or more) the density of cells within the forward separation

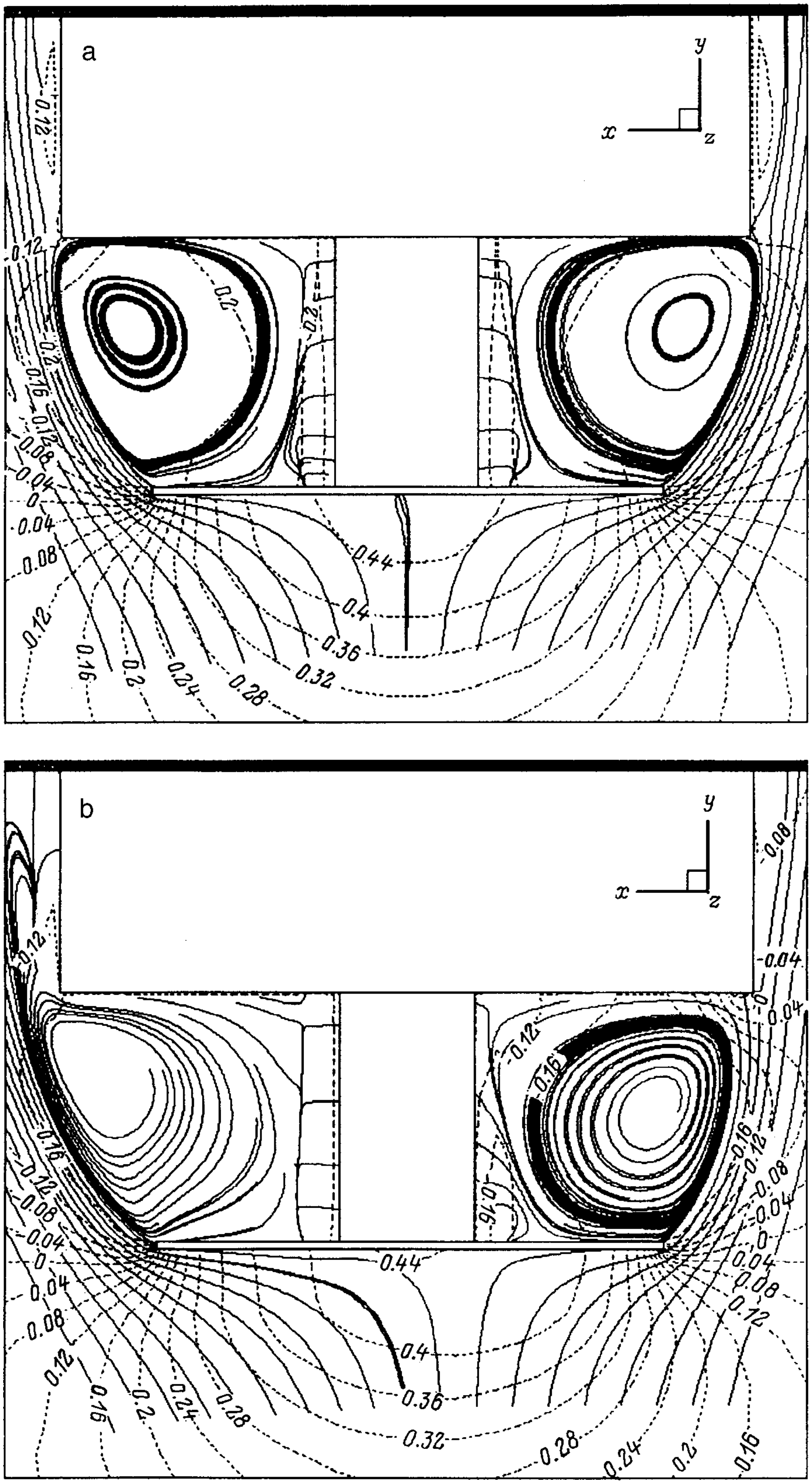

FIG. 1. Pattern of flow around a semi-infinite cylinder with a disk of diameter $d=0.75$ and gap $l=0.375$ for Reynolds number $\mathrm{Re}=10^3$ in the axial $x-y$ plane at angles of attack of 0 (a) and 5° (b). The dashed lines show isobars (the excess pressure is referenced to a value of twice the velocity-related pressure head).

zone. Furthermore, in order to avoid problems connected with interpolation of the metric coefficients, we use a simplified approach based on prespecifying an analytic metric. In all other details the methodology of our investigation does not differ from that described in Ref. 4. By using an implicit finite-volume method and the concept of splitting the problem with respect to physical processes, we solved the system of Navier–Stokes equations given in the natural variables for the Cartesian components of the velocity. A uniform flow was specified at the input boundary of the computational region. At the output boundary we imposed soft boundary conditions (the conditions of continuation of the solution

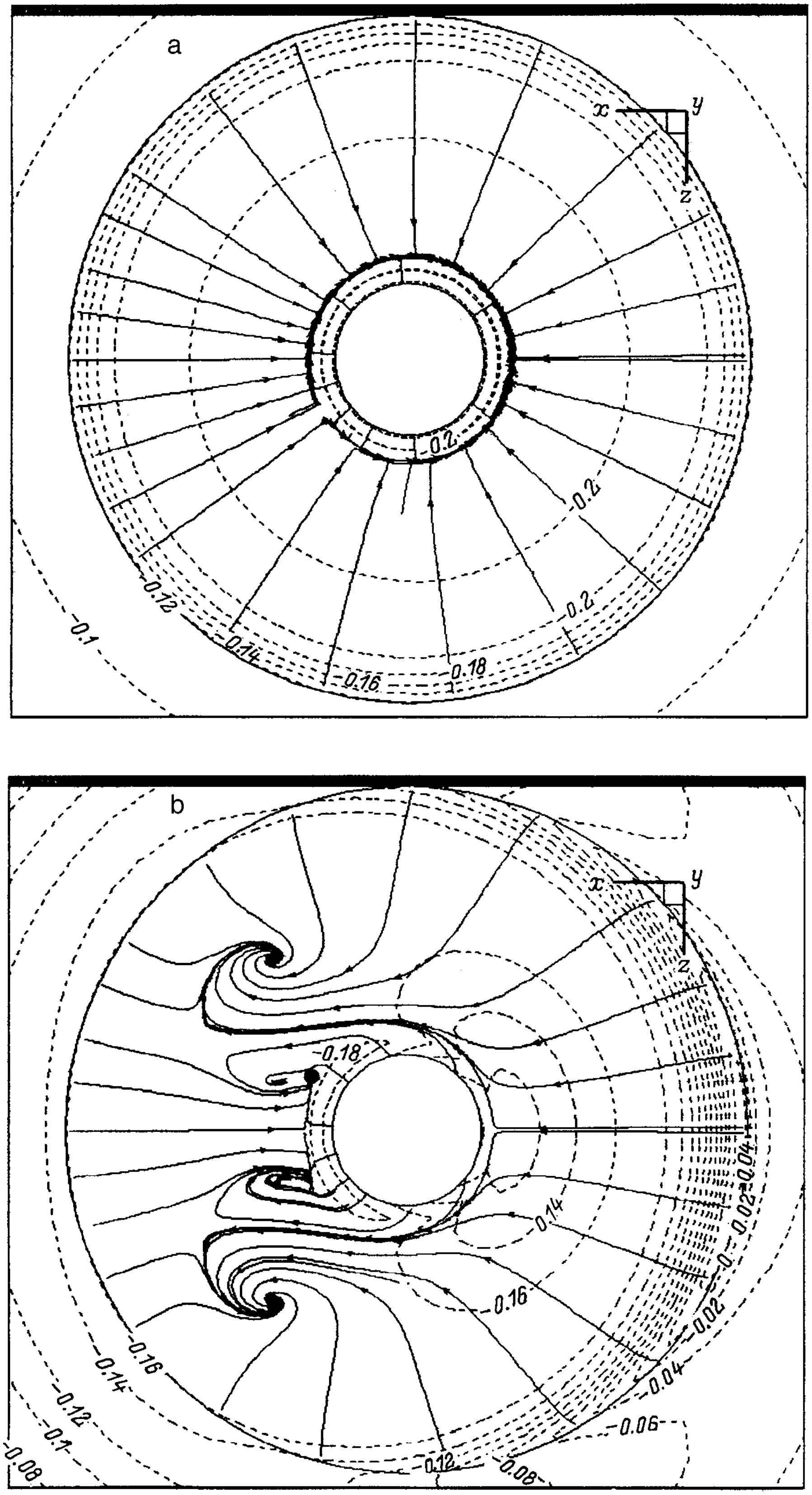

FIG. 2. Spreading patterns of fluid over the end of a cylinder with a projecting disk ($\mathrm{Re}=10^3$) at angles of attack of 0 (a) and 5° (b). The dotted lines show isobars.

from internal points to the boundary of the region). At the solid surfaces the condition of attachment was imposed. The steady flow around the body was calculated on a mesh containing $51\times 66\times 61$ cells. In this case, approximately 30 000 points were concentrated in the gap between the disk and the cylinder.

In Figures 1 and 2 and in Table I we systematically compare our calculated results for laminar flow past a semi-

TABLE I. Combined aerodynamic characteristics of the blunt bodies discussed here and of their separate components.

	Semi-infinite cylinder		Disk–semi-infinite cylinder arrangement with $d=0.75$, $l=0.375$			
Re	700		700		1000	
α	0	5°	0	5°	0	5°
C_{X1}	0.762	0.779	-0.212	-0.184	-0.367	-0.293
C_{X2}	–	–	0.42	0.42	0.378	0.367
C_{X3}	–	–	0.142	0.128	0.211	0.178
C_{X4}	0.762	0.779	0.349	0.363	0.221	0. 252
M_{Z1}	-0.506×10^{-4}	-0.913×10^{-2}	0.183×10^{-3}	-0.0154	0.176×10^{-3}	-0.0124
M_{Z2}	–	–	0.28×10^{-4}	-0.00382	-0.302×10^{-4}	-0.00495
M_{Z3}	–	–	-0.266×10^{-4}	0.00197	-0.201×10^{-4}	0.482×10^{-3}
M_{Z4}	-0.506×10^{-4}	-0.913×10^{-2}	-0.184×10^{-3}	-0.0167	0.126×10^{-3}	-0.0168
V_m	-0.0606	-0.102	-0.146	-0.31	-0.192	-0.27

Note: C_X is the drag coefficient, M_Z is the transverse moment coefficient. Labels 1–4 refer to the end of the cylinder, the forward face of the disk, the backward face of the disk, and the assembly type respectively; m denotes minimum value.

infinite cylinder with (for Re = 700 and 1000) and without (for Re = 700) a disk in front of it, at angles of attack of 0 and 5°. Note that the results obtained correlate with the solutions obtained in Ref. 4 and 5 for axisymmetric laminar flow past a cylinder with a disk, and with visualization data for flow separation in a model tank.

Placing a disk in front of the cylinder leads to a considerable rearrangement of the flow around it, due to the formation of a forward separation zone and hence to a decrease in the profile drag (C_{X4}) by a factor of two. It is clear from Fig. 1a that toroidal vortices appear in the gap between the disk and the cylinder at zero angle of attack, in which the intensity of the return flow (V_m) is 15% of the velocity of the oncoming flow. The low (negative) pressure in the vortex, which is practically the same for the entire forward separation zone (Fig. 2a), gives rise to a thrust exerted on the cylinder (C_{X1}), which is the ultimate source of the lowered drag of blunt bodies with a projecting disk. With increasing Reynolds number the intensity of flow separation increases, and the profile drag falls considerably (by more than a factor of 3 compared with C_{X4} for a cylinder).

For a nonzero (but comparatively low) angle of attack, the flow past a cylinder with a disk acquires a very complicated spatial character. The toroidal vortex is distorted, especially in the axial plane parallel to the oncoming flux. In this case the vortex becomes markedly nonuniform. The intensification of return flow from the windward region (in which V_m grows by a factor of 2, reaching 30% of the velocity of the oncoming flow) combines with a weakening of the separated flow on the leeward side. As a consequence of the distortion of the vortex, a redistribution of the local loading forces takes place along the back end of the cylinder, giving rise to a considerable restoring torque M_{Z4} (i.e., bow stabilization).

The complicated vortex structure of the flow is characterized by the appearance of jet streams at singular runoff points of the ‘‘focus’’ type at the end of the cylinder (Fig. 2b) and a reflux of fluid from the windward side of the toroidal vortex to the leeward side and then onto the lateral surface, with suction of the liquid from the leeward side to the windward side.

This work was carried out with the financial support of the Russian Fund for Fundamental Research, Projects Nos. 95-01-00728 and 96-01-00298.

[1] I. A. Belov, *Interaction of Nonuniform Flows with Obstacles* [in Russian] (Mashinostroenie, Leningrad, 1983), 144 pp.

[2] I. A. Belov and S. A. Isaev, Pis'ma Zh. Tekh. Fiz. **6**, 608 (1980) [Sov. Tech. Phys. Lett. **6**, 264 (1980)].

[3] K. Koenig and A. Roshko, J. Fluid Mech. **156**, 167 (1985).

[4] I. A. Belov, S. A. Isaev, and V. A. Korobkov, *Problems and Computational Methods for Separated Flows of Incompressible Fluids* [in Russian] (Sudostroenie, Leningrad, 1989), 256 pp.

[5] V. K. Bobyshev, S. A. Isaev, and O. L. Lemko, Inzh.-Fiz. Zh. **51**, 224 (1986).

[6] V. K. Bobyshev and S. A. Isaev, Inzh.-Fiz. Zh. **58**, 556 (1990).

[7] S. A. Isaev, Inzh.-Fiz. Zh. **68**, 19 (1995).

[8] I. A. Belov, V. K. Bobyshev, and S. A. Isaev, in *Investigations of the Aerodynamics and Dynamics of Flight* [in Russian] (KIIGA, Kiev, 1988), pp. 122–129.

[9] S. A. Isaev and N. A. Sumovskiĭ, Inzh.-Fiz. Zh. **70**, 990 (1997).

Translated by Frank J. Crowne