THEORETICAL
AND MATHEMATICAL PHYSICS

Thermoelectric Phenomena at Grain Boundaries

A. V. Latyshev and A. A. Yushkanov

Moscow State Regional University, ul. Radio 10a, Moscow, 105005 Russia
e-mail: latyshev@comail.ru, yushkanov@mtu-net.ru
Received February 24, 2004

Abstract—Analytical expressions for a temperature jump and electric potential difference that arise when current passes through a grain boundary are derived. The electron flow (current) through the boundary and the current-induced heat flux are assumed to be given. The kinetic equation in the τ approximation for electrons and the Maxwell equation for an electric field are used. The dependence of the temperature jump and potential difference factors on the chemical potential is studied. © 2004 MAIK "Nauka/Interperiodica".

INTRODUCTION

Thermoelectric phenomena at the metal–metal interface have been the subject of extensive research for a long time [1, 2]. Classical thermoelectric phenomena, such as Peltier and Thomson effects, are finding wide application. In addition, they are of scientific interest. Note that classical thermoelectric phenomena arise when the surfaces of different metals are in contact.

However, there also exist thermoelectric phenomena arising when identical metals come into contact. Examples are temperature and potential jumps that occur when current or a heat flux passes through the metal–metal interface. These jumps are observed at the contact of both different and identical metals. The latter case, where the interface is, in essence, a grain boundary in a polycrystalline metal, is considered in this work.

Most metals are polycrystals; therefore, the estimation of the temperature and/or potential jumps when current crosses a grain boundary is of great significance.

We assume that the current is directed normally to the interface and the current value and corresponding heat flux are given. The electron behavior will be described in terms of the kinetic equation in the τ approximation; the behavior of the electric field, in terms of the Maxwell equation. Modifying our approach developed in [3, 4], we will find an exact solution to this set of equations, as well as the values of the temperature and potential jumps as linear functions of the current and heat flux. The coefficients of these two parameters are chemical-potential-dependent functions.

Let the Fermi surface for a metal considered be spherical and the electric field be directed normally to a grain boundary selected. The electric field is assumed to be low enough, so that the linear approximation is valid [1, 2].

Let us direct the x axis normally to the surface and place the origin on the boundary. It is assumed that a heat flux passes in the metal along the x axis. Then, at distances much greater than electron mean free path λ, the heat flux can be represented by constant temperature gradient $G_T = dT/dx$ (the metal is assumed to be isotropic). The temperature gradient is assumed to be flat: the relative temperature drop over length λ is much smaller than unity.

The layer of thickness λ that is adjacent to the boundary will be referred to as the Knudsen layer, as is customary in kinetic considerations. Outside the Knudsen layer, the temperature profile has the form $T = T_{0+} + G_T x$ for $x > 0$ and $T = T_{0-} + G_T x$ for $x < 0$. The value of $\Delta T = T_{0-} - T_{0+}$ will be called the temperature jump. Because of the linear statement of the problem, $\Delta T = C_T \lambda G_T$, where C_T independent coefficient C_T will be called the temperature jump coefficient. Sometimes, it is more convenient to handle dimensionless temperature jump $\varepsilon_T = \Delta T/T_s$, where T_s is the temperature of the surface. Then, $\varepsilon_T = C_T \lambda g_T$, where $g_T = G_T/T_s$ is the relative temperature gradient.

Similarly, when the electric field is normal to the surface, the profile of potential U in the metal has the form $U = U_{0+} - E_0 x$ for $x > 0$ and $U = U_{0-} - E_0 x$ for $x < 0$. Here, E_0 is the electric field strength away from the boundary. Near the boundary, the field becomes variable and, accordingly, the profile becomes nonlinear. The value of $\Delta U = U_{0-} - U_{0+}$ will be called the potential jump. In this case, too, we have $\Delta U = C_U \lambda G_T$, where coefficient C_U is the potential jump coefficient, because of the linear statement of the problem.

In metals, heat transfer usually generates an electric field. Therefore, the problem of temperature distribution near a metal surface is often solved in parallel with the problem of electric field in the Knudsen layer, which is associated with thermal processes.

The aim of our work is (i) to calculate the temperature jump in metals using an analytical solution to the

kinetic equation for electrons, (ii) find the electric potential difference near the surface, and (iii) construct the electric field profile.

We consider the case when the electron gas is arbitrarily degenerate. Therefore, the results obtained in this work apply to a wide temperature range and a variety of materials (including semimetals).

KINETIC EQUATION AND PROBLEM DEFINITION

For electrons in metals, the kinetic equation in the τ approximation [1, 2] is frequently used:

$$\frac{\partial f}{\partial t} + (\mathbf{v} \cdot \nabla) f + e_0 \mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{p}} = \frac{1}{\tau}(f_F^0 - f). \qquad (1)$$

Here, f is the electron distribution function, e_0 is the electron charge, $\mathbf{p}$ is the electron momentum, $\mathbf{E}$ is the electric field, $\mathbf{v}$ is the electron velocity, τ is the electron relaxation time, and f_F^0 is the Fermi distribution function. If the temperature is finite, Eq. (1) must involve Fermi distribution function f_F^* with some effective temperature T_* and effective chemical potential μ_* as the equilibrium electron distribution function instead of f_F^0, which corresponds to the zero temperature. Then, Eq. (1) takes the form

$$\frac{\partial f}{\partial t} + (\mathbf{v} \cdot \nabla) f + e_0 \mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{p}} = \frac{v}{\lambda}(f_F^* - f). \qquad (2)$$

Here, $f_F^* = f_F(\mu_*, T_*) = \left[\exp\left(\frac{mv^2}{2kT_*} - \frac{\mu_*}{kT_*}\right) + 1\right]^{-1}$ is the Fermi distribution function (Fermian), k is the Boltzmann constant, and m is the electron effective mass (for details, see [3]).

In most metals, the electron subsystem makes a major contribution to heat transfer [5]. We will consider just this case and ignore the phonon contribution to this process. Also, we assume that the mass velocity of the electron gas is much lower than the electron velocity and that characteristic temperature differences over length λ are much smaller than the electron gas temperature. These assumptions allow us to linearize the problem.

Let us introduce the designations

$$\mathbf{c} = \sqrt{\beta_s} \mathbf{v}, \quad \beta_s = \frac{m}{2kT_s}, \quad \alpha = \frac{\mu_s}{kT_s},$$

$$\varepsilon_* = \frac{mv^2}{2kT_*} - \frac{\mu_*}{kT_*},$$

as well as dimensionless time $t_* = \sqrt{T_s/\beta_s}\, t/\lambda$ and coordinate $\mathbf{r}_* = \mathbf{r}/\lambda$, and put $g_{T_*} = \lambda g_T$. Now, we linearize Eq. (2). Applying the particle and energy conservation

laws, we arrive at the equation [3] (the asterisks by the variables and gradient will hereafter be omitted)

$$\frac{\partial \varphi}{\partial t} = c\varphi(t, \mathbf{r}, \mathbf{c}) - \mathbf{c} \cdot \mathbf{e}(\mathbf{r})$$

$$= \frac{c}{2\pi}\int k(c, c')\varphi(t, \mathbf{r}, \mathbf{c}')d\Omega(\alpha). \qquad (3)$$

Here, function φ is related to distribution function f via the relationship $f = f_F^s + g\varphi$, where $f_F^s = f_F(\mu_s, T_s)$;

$$k(c, c') = 1 + \frac{g_1^2(\alpha)}{\Delta(\alpha)}\left(c^2 - \frac{g_3(\alpha)}{g_1(\alpha)}\right)\left(c'^2 - \frac{g_3(\alpha)}{g_1(\alpha)}\right),$$

$$g(c) = g(c, \alpha) = \frac{\exp(c^2 - \alpha)}{[1 + \exp(c^2 - \alpha)]^2},$$

$$d\Omega(\alpha) = \frac{g(c', \alpha)}{g_1(\alpha)}d^3c',$$

$$g_3(\alpha) = 2\int_0^\infty g(c, \alpha)c^5 dc = 4\int_0^\infty c\ln[1 + \exp(\alpha - c^2)]c\,d,$$

$$g_1(\alpha) = 2\int_0^\infty g(c, \alpha)c^3 dc = \ln(1 + \exp(\alpha)),$$

$$g_5(\alpha) = 2\int_0^\infty g(c, \alpha)c^7 dc$$

$$= 12\int_0^\infty c^3\ln[1 + \exp(\alpha - c^2)]dc,$$

$$\Delta(\alpha) = g_1(\alpha)g_5(\alpha) - g_3^2(\alpha), \quad \mathbf{e}(\mathbf{r}) = \frac{e_0\lambda}{kT_s}\mathbf{E}(\mathbf{r}).$$

Suppose that the reflection of the electrons from the boundary is perfectly diffuse. Then, the set of equations that describe the problem consists of Eq. (3) and the equation for the electric field. In dimensionless form, they are written as

$$\mu\frac{\partial \varphi}{\partial x} + \varphi(x, \mu, c) - \mu e(x)$$

$$= \frac{1}{g_1(\alpha)}\int_{-1}^1\int_0^\infty k(c, c')\varphi(x, \mu', c')g(c')c'^3 d\mu'dc', \qquad (4)$$

$$e'(x) = a_0^2\int_{-1}^1\int_0^\infty \varphi(x, \mu, c)g(c)c^2 d\mu dc, \qquad (5)$$

$$a_0^2 = \frac{e_0^2 m^2 \lambda^2}{\pi^2 \varepsilon_0 \hbar^3 \sqrt{\beta_s}}, \quad \mu = \frac{c_x}{c},$$

where ε_0 is the dielectric constant.

The boundary conditions and the conditions inside the metal for $x > 0$ are as follows:

$$\varphi(+0, \mu, c) \equiv \varphi_0(\mu, c) = A_0, \quad 0 < \mu < 1,$$

$$\varphi(x, \mu, c) = \varphi_{as}^+(x, \mu, c) + o(1), \quad x \longrightarrow +\infty,$$

$$e(\infty) = e_{as}.$$

Accordingly, for $x < 0$, we have

$$\varphi(-0, \mu, c) = -A_0, \quad -1 < \mu < 0,$$

$$\varphi(x, \mu, c) = \varphi_{as}^-(x, \mu, c) + o(1), \quad x \longrightarrow -\infty,$$

$$e(-\infty) = e_{as}.$$

Here,

$$\varphi_{as}^+(x, \mu, c) = e_{as}\mu + [\varepsilon_T^+ + g_T(x-\mu)]\left(c^2 - \frac{g_2(\alpha)}{s(\alpha)}\right),$$

$$\varepsilon_T^+ = \frac{T_{0+} - T_s}{T_s},$$

$$\varphi_{as}^-(x, \mu, c) = e_{as}\mu + [-\varepsilon_T^+ + g_T(x-\mu)]\left(c^2 - \frac{g_2(\alpha)}{s(\alpha)}\right),$$

$$g_2(\alpha) = 2\int_0^\infty g(c, \alpha)c^2 dc = \frac{3}{2}\int_0^\infty \ln[1 + \exp(\alpha - c^2)]dc,$$

$$g_0(\alpha) = 2\int_0^\infty g(c, \alpha)c^2 dc = \int_0^\infty \frac{\exp(\alpha - c^2)}{1 + \exp(\alpha - c^2)}dc.$$

We took advantage of the fact that $(T_{0+} - T_s)/T_s = -(T_{0-} - T_s)/T_s$ owing to the symmetry of the problem.

BALANCE OF ELECTRON FLOWS ON THE SURFACE

Let

$$N_0 = \int_{c_x > 0} \varphi(0, \mu, c)g(c)(\mu c)d^3c$$

be the electron flow that strikes the grain boundary from the left-hand half-space,

$$N_1 = \int_{c_x > 0} A_0 g(c)(\mu c)d^3c = \frac{\pi}{2}g_1(\alpha)A_0$$

be the electron flow crossing the boundary from left to right, and

$$N_2 = \int_{c_x < 0} \varphi(0, \mu, c)g(c)(\mu c)d^3c$$

be the electron flow striking the boundary from the right-hand half-space.

The flow balance equation has the form

$$N_1 = pN_0 - (1 - p)N_2,$$

where p is the probability of electron scattering in a forward direction. Also, heat flux Q_0 equals heat flux Q_2 ($Q_0 = Q_2$).

In view of this equality, $N_1 = 2(p - 1/2)N_2$. Hence,

$$A_0 = \frac{(4p - 2)}{\pi g_1(\alpha)}\int_{c_x < 0} \varphi(0, \mu, c)g(c)\mu c d^3c$$

or

$$A_0 = \frac{8p - 4}{g_1(\alpha)}\int_{-1}^0 \mu d\mu \int_0^\infty \varphi(0, \mu, c)g(c)c^3 dc.$$

For isotropic scattering ($p = 0.5$), $A_0 = 0$. The expression for A_0 can be represented as

$$A_0 = \frac{8p - 4}{g_1(\alpha)}\left[\int_{-1}^1 \mu d\mu \int_0^\infty \varphi(0, \mu, c)g(c)c^3 dc \right.$$
$$\left. - A_0 \int_0^1 \mu d\mu \int_0^\infty g(c)c^3 dc\right].$$

Then, it follows that

$$A_0 = \frac{8p - 4}{g_1(\alpha)}\int_{-1}^0\int_0^\infty \varphi(0, \mu, c)\mu c^3 g(c)dc - (2p - 1)A_0.$$

The integral term in this equality is the integral electron flow. This quantity is constant by virtue of the electron (charge) conservation law. Therefore, we replace function $\varphi(0, \mu, c)$ by $\varphi_{as}(0, \mu, c)$ is the asymptotic value) and, calculating

$$A_0 = \frac{4p - 2}{pg_1(\alpha)}\int_{-1}^1\int_0^\infty \left[e_{es}\mu + (\varepsilon_T - g_T\mu)\left(c^2 - \frac{g_2(\alpha)}{g_0(\alpha)}\right)\right]$$
$$\times g(c)\mu c^3 d\mu dc,$$

obtain

$$A_0 = \frac{4p - 2}{3p}(e_{as} + h_0(\alpha)g_T),$$

where

$$h_0(\alpha) = \frac{\Delta_0(\alpha)}{g_0(\alpha)g_1(\alpha)},$$

$$\Delta_0(\alpha) = g_2(\alpha)g_1(\alpha) - g_3(\alpha)g_0(\alpha).$$

ANALYTICAL SOLUTION
OF THE PROBLEM

The kinetic equation and the structure of the boundary conditions allow us to consider the half-space $x > 0$ alone (the case $x < 0$ is treated in a similar way). Taking into account the structure of function φ_{as}, we seek function φ in the form

$$\varphi(x, \mu, c) = h_1(x, \mu) + \left(c^2 - \frac{g_2(\alpha)}{g_0(\alpha)}\right)h_2(x, \mu).$$

Eventually, we arrive at the following subproblems. For function $h_1(x, \mu)$:

$$\mu\frac{\partial h_1}{\partial x} + h_1(x, \mu) = \frac{1}{2}\int_{-1}^{1} h_1(x, \mu')d\mu' + \mu e(x),$$

$$h_1(0, \mu) = A_0, \quad 0 < \mu < 1,$$

$$h_1(x, \mu) = e_{as}\mu + o(1), \quad x \longrightarrow +\infty, \quad -1 < \mu < 0;$$

For function $h_2(x, \mu)$:

$$\mu\frac{\partial h_2}{\partial x} + h_2(x, \mu) = \frac{1}{2}\int_{-1}^{1} h_2(x, \mu')d\mu',$$

$$h_2(0, \mu) = 0, \quad 0 < \mu < 1,$$

$$h_2(x, \mu) = \varepsilon_T^+ + G_T(x - \mu) + o(1),$$

$$x \longrightarrow +\infty, \quad -1 < \mu < 0.$$

For the electric field, we get

$$e'(x) = a^2\frac{1}{2}\int_{-1}^{1} h_1(x, \mu)d\mu, \quad e(\infty) = e_{as},$$

$$a = a_0\sqrt{g_0(\alpha)}.$$

Consider the subproblem for $h_2(x, \mu)$, which is the problem of temperature jump. This problem was solved in [3]. The temperature jump is found from the formula

$$\Delta T^+ = T_{0+} - T_s = V_1 G_T,$$

$$V_1 = -\frac{1}{\pi}\int_{0}^{1} \zeta(\tau)d\tau = 0.71045,$$

$$\zeta(\tau) = -\frac{\pi}{2} - \arctan\left(\frac{2}{\pi\tau} + \frac{1}{\pi}\ln\frac{1-\tau}{1+\tau}\right). \quad (6)$$

The value of the temperature jump is the same on both sides of the grain boundary: $\Delta T^- = T_{0-} - T_s = -\Delta T^+$. Therefore, the total temperature jump at the boundary is twice as large as the partial jump. Hence,

$$\Delta T = \Delta T^- - \Delta T^+ = -2V_1 G_T. \quad (7)$$

According to [3, 4], functions $h_1(x, \mu)$ and $e(x)$ are sought in the form of expansions:

$$h_1(x, \mu) = e_{as}\mu + A_1\exp(-ax)$$
$$+ \int_{0}^{1} \exp\left(-\frac{x}{\eta}\right)F(\eta, \mu)n(\eta)d\eta, \quad (8)$$

$$e(x) = e_{as} - aA_1\exp(-ax) - \frac{a^2}{2}\int_{0}^{1} \exp\left(-\frac{x}{\eta}\right)n(\eta)d\eta. \quad (9)$$

Here, A_1 is an unknown constant, $n(\eta)$ is an unknown function,

$$F(\eta, \mu)$$
$$= \frac{a^2}{2}\eta + \frac{1 - a^2\eta^2}{\eta}\left[\frac{\eta}{2}P\frac{1}{\eta - \mu} + \lambda(\eta)\delta(\eta - \mu)\right],$$

Px^{-1} stands for the principal value of the integral of x^{-1}, $\delta(x)$ is the Dirac function, and

$$\lambda(z) = 1 + \frac{z}{2}\int_{-1}^{1} \frac{d\tau}{\tau - z},$$

is the Case dispersion function [6].

Substituting expansions (8) and (9) into the corresponding boundary conditions yields

$$A_0 = e_{es}\mu + A_1 + A_2$$
$$+ \frac{1}{2}\int_{0}^{1} (1 - a^2\eta^2)\frac{n(\eta)d\eta}{\eta - \mu} + (1 - a^2\mu^2)\frac{\lambda(\mu)}{\mu} = 0, \quad (10)$$

$$0 < \mu < 1, \quad A_2 = \frac{a^2}{2}\int_{0}^{1} \eta n(\eta)d\eta$$

and

$$e_{as} - e(0) - aA_1 - \frac{a^2}{2}\int_{0}^{1} n(\eta)d\eta = 0. \quad (11)$$

Solving Eq. (10), we find

$$A_1 = A_0 - A_2 - e_{as}V_1,$$

$$(1 - a^2\eta^2)n(\eta) = \frac{e_{as}}{\pi i}\left[\frac{1}{X^+(\eta)} - \frac{1}{X^-(\eta)}\right] = -\frac{2e_{as}}{\pi}\frac{\sin\zeta(\eta)}{X(\eta)}, \quad (12)$$

$$aA_1 = -e(0) + \frac{a}{2}e_{as}\left[\frac{\cos\zeta(1/a)}{X(1/a)} - \frac{1}{X(-1/a)}\right],$$

$$A_2 = -e_{as}\left[V_1 + \frac{\cos\zeta(1/a)}{2X(1/a)} + \frac{1}{2X(-1/a)}\right],$$

where

$$X(z) = \frac{1}{z} \exp V(z), \quad V(z) = \frac{1}{\pi} \int\limits_0^1 \frac{\zeta(\tau) d\tau}{\tau - z}.$$

Substituting the found values of A_0, A_1, and A_2 into (12), we find the electric field at the grain boundary:

$$e(0) = -\frac{a(4p-2)}{3p}(e_{as} + h_0(\alpha)g_T) - \frac{ae_{as}}{X(-1/a)}. \quad (13)$$

POTENTIAL AND TEMPERATURE JUMPS, CURRENT, AND HEAT FLUX

By definition, the potential jump is found by the formula

$$\Delta U = \int\limits_{-\infty}^{\infty} [e(x) - e_{as}]dx = 2\int\limits_0^{\infty} [e(x) - e_{as}]dx$$

$$= -2A_1 - a^2 \int\limits_0^1 \eta n(\eta) d\eta = -2(A_1 + A_2)$$

or, according to (12),

$$\Delta U = -2(A_0 - V_1 e_{as})$$

$$= -e_{as}\left(\frac{8p-4}{3p} - 2V_1\right) - \frac{8p-4}{3p}h_0(\alpha)g_T. \quad (14)$$

In terms of the distribution function, current density j is expressed

$$j = 2e_0\left(\frac{m}{2\pi\hbar}\right)^3 \int v_x \varphi(x, \mathbf{v}, t) d^3 v.$$

Passing to the dimensionless variables and substituting φ_{as} for φ, we have $(e_0 < 0)$

$$j = -\frac{1}{\delta_j}[g_1(\alpha)e_{as} + h_0(\alpha)g_1(\alpha)g_T],$$

$$\delta_j = -\frac{24\pi^2\hbar^3 T_s^2 k^2}{e_0 m^5} > 0. \quad (15)$$

The quantity

$$\sigma = -\frac{e_0\lambda}{kT_s\delta_j}g_1(\alpha) > 0$$

equals the electric conductivity of the metal.

The heat flux in the metal is calculated through function f by the formula

$$Q = 2\int \frac{mv^2}{2}v_x f \frac{d^3 p}{(2\pi\hbar)^3}.$$

Now, passing to the dimensionless variables and using the relationship $f = f_F^s + \varphi g$, we get

$$Q = \frac{1}{\delta_Q}[g_3(\alpha)e_{as} + g_3(\alpha)h_1(\alpha)g_T], \quad \delta_Q = \frac{\pi^2\hbar^3}{mk^3 T_s^3}, \quad (16)$$

where

$$h_1(\alpha) = \frac{\Delta_1(\alpha)}{g_3(\alpha)g_0(\alpha)},$$

$$\Delta_1(\alpha) = g_2(\alpha)g_3(\alpha) - g_0(\alpha)g_5(\alpha)$$

and the quantity

$$\chi = -\frac{\lambda}{T_s\delta_Q}\frac{D_1(\alpha)}{g_0(\alpha)} > 0$$

is the thermal conductivity of the metal.

Let us express the temperature and potential jumps through the current and heat flux. From (15) and (16), we find

$$g_T = -\frac{1}{\Delta(\alpha)}[j\delta_i g_3(\alpha) + Q\delta_Q g_1(\alpha)],$$

$$e_{as} = \frac{1}{g_0(\alpha)\Delta(\alpha)}[j\delta_i\Delta_1(\alpha) + Q\delta_Q\Delta_0(\alpha)].$$

Now, we find the temperature jump (in dimensional form) and potential jump from Eqs. (7) and (14) in view of (17) and (18):

$$\Delta T = \frac{2V_1 T_s}{\Delta(\alpha)}[j\delta_j g_3(\alpha) + Q\delta_Q g_1(\alpha)],$$

$$\Delta U = j\delta_j\left[\frac{8p-4}{3pg_1(\alpha)} + 2V_1\frac{\Delta_1(\alpha)}{g_0(\alpha)\Delta(\alpha)}\right] + Q\delta_Q\frac{2V_1\Delta_0(\alpha)}{g_0(\alpha)\Delta(\alpha)}.$$

For the potential jump in dimensional form, we have

$$\Delta U = 2\int\limits_0^{\infty}(E(x) - E_{as})dx,$$

where $x = \lambda x'$ is the dimensional coordinate and x' is the dimensionless coordinate.

Since

$$E(x) = \frac{kT_s}{e_0\lambda}e(x),$$

the dimensional potential jump takes the form

$$\Delta U = \frac{kT_s\delta_j}{\lambda e_0}\left[\frac{8p-4}{3pg_1(\alpha)} + 2V_1\frac{\Delta_1(\alpha)}{g_0(\alpha)\Delta(\alpha)}\right]j$$

$$+ \frac{kT_s\delta_Q}{\lambda e_0}\frac{2V_1\Delta_0(\alpha)}{g_0(\alpha)\Delta(\alpha)}Q.$$
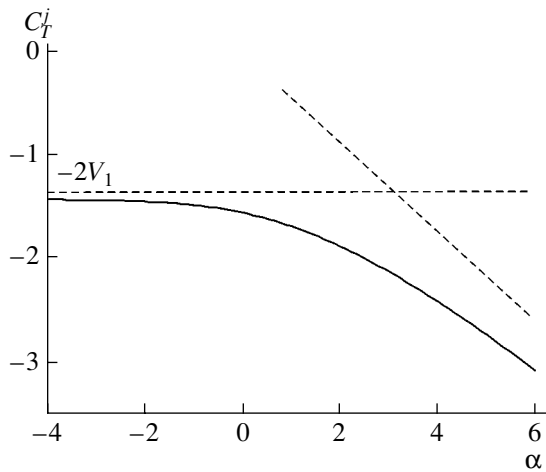
Let us introduce dimensionless kinetic coefficients
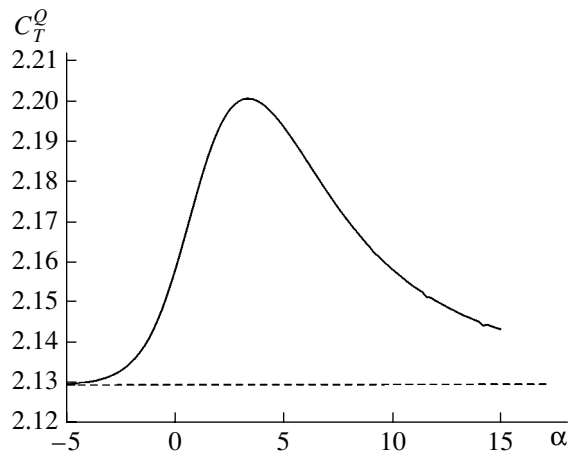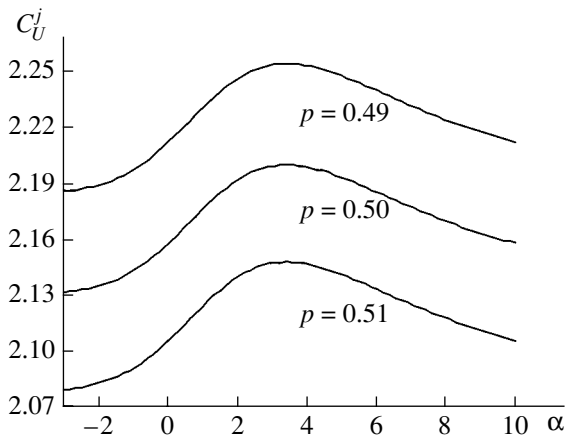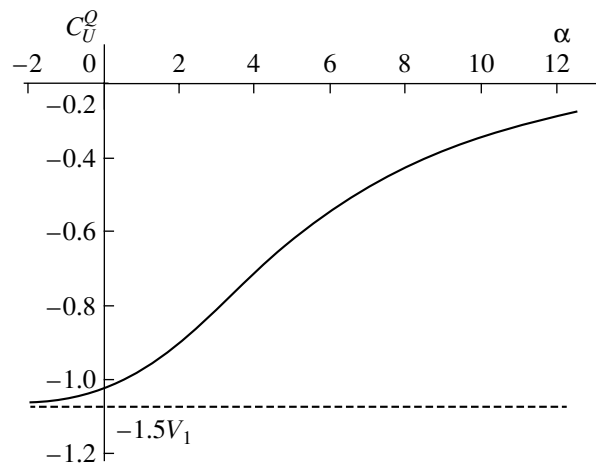
Fig. 1.

Fig. 2.

Fig. 3.

Fig. 4.

(Figs. 1–4)

$$C_T^j(\alpha) = -2V_1 \frac{g_1(\alpha)g_3(\alpha)}{\Delta(\alpha)},$$

$$C_T^Q(\alpha) = -2V_1 \frac{g_1(\alpha)\Delta_1(\alpha)}{g_0(\alpha)\Delta(\alpha)},$$

$$C_U^j(\alpha, p) = -\left[\frac{8p-4}{3p} + 2V_1 \frac{g_1(\alpha)\Delta_1(\alpha)}{g_0(\alpha)\Delta(\alpha)}\right],$$

$$C_U^Q(\alpha) = -2V_1 \frac{\Delta_0(\alpha)\Delta_1(\alpha)}{g_0^2(\alpha)\Delta(\alpha)}.$$

In terms of the electric conductivity and thermal conductivity of the metal (see above), the temperature and potential jumps can be expressed in dimensional form as follows:

$$\Delta T = \frac{e_0\lambda}{\sigma k}C_T^j(\alpha)j + \frac{\lambda}{\kappa}C_T^Q(\alpha)Q, \qquad (17)$$

$$\Delta U = \frac{1}{\sigma}C_U^j(\alpha, p)j - \frac{k}{e_0\kappa}C_U^Q(\alpha)Q. \qquad (18)$$

Thus, the heat flux (Q) causes a temperature jump along with a potential jump, while the current causes a potential jump along with a temperature jump; that is, cross effects take place.

Of the coefficients involved in formula (17), coefficient C_U^j alone depends on probability p of electron scattering in a forward direction.

Note in conclusion that quantity $\rho_s = C_U^j/\sigma$ may be viewed as the grain boundary electrical resistivity (the resistance per unit area). Similarly, quantity $\lambda C_Q^T/\kappa$ may be viewed as the grain boundary specific thermal resistance. The value of ρ_s indefinitely grows as probability p of forward electron scattering tends to zero, since the grain boundary becomes nontransparent to electrons in this case.

LIMITING CASES

Let us find the asymptotics of the kinetic coefficients at $\alpha \longrightarrow +\infty$ (degenerate plasma) and $\alpha \longrightarrow -\infty$ (conventional plasma). For the first case, we take advantage of formula (58.1) from [7, p. 191]:

$$\int_0^\infty \frac{f(x)dx}{\exp(x-\alpha)+1} = \int_0^a f(x)dx + \frac{\pi^2}{6}f'(\alpha)$$
$$+ \frac{7\pi^2}{360}f'''(\alpha) + \ldots, \quad \alpha \gg 1.$$

Based on this formula, we have for $\alpha \longrightarrow +\infty$

$$g_0(\alpha) = \alpha^{1/2} - \frac{\pi^2}{24}\alpha^{-3/2} + \ldots, \quad g_1(\alpha) = \alpha + \ldots,$$

$$g_2(\alpha) = \alpha^{3/2} + \frac{\pi^2}{8}\alpha^{-1/2} + \ldots, \quad g_3(\alpha) = \alpha^2 + \frac{\pi^2}{3} + \ldots,$$

$$g_4(\alpha) = \alpha^{5/2} + \frac{5\pi^2}{8}\alpha^{1/2} + \ldots, \quad g_5(\alpha) = \alpha^3 + \pi^2\alpha + \ldots.$$

Hence, at $\alpha \longrightarrow +\infty$,

$$\Delta_0(\alpha) = -\frac{\pi^2}{6}\alpha^{1/2} + \ldots, \quad \Delta_1(\alpha) = -\frac{\pi^2}{2}\alpha^{3/2} + \ldots,$$

$$\Delta(\alpha) = \frac{\pi^2}{3}\alpha^2 + \ldots.$$

Consequently, the asymptotics of the kinetic coefficients at $\alpha \longrightarrow +\infty$ (degenerate plasma) is as follows:

$$C_T^j(\alpha) = -\frac{6V_1}{\pi^2}\alpha + \ldots, \quad C_T^Q(\alpha) = 3V_1 + \ldots,$$

$$C_U^j(p, \alpha) = 3V_1 - \frac{8p-4}{3p} + \ldots, \quad C_Q^U(\alpha) = -\frac{V_1\pi^2}{\alpha} + \ldots.$$

For the case of classical gas at $\alpha \longrightarrow -\infty$, we have

$$g_0(\alpha) = \frac{\sqrt{\pi}}{2}\exp\alpha, \quad g_1(\alpha) = \exp\alpha,$$

$$g_2(\alpha) = \frac{3\sqrt{\pi}}{4}\exp\alpha,$$

$$g_3(\alpha) = 2\exp\alpha, \quad g_4(\alpha) = \frac{15}{8}\sqrt{\pi}\exp\alpha,$$

$$g_5(\alpha) = 6\exp\alpha,$$

$$\Delta\alpha = 2\exp(2\alpha), \quad \Delta_0(\alpha) = -\frac{\sqrt{\pi}}{4}\exp(2\alpha),$$

$$\Delta_1(\alpha) = -\frac{3}{2}\sqrt{\pi}\exp(2\alpha).$$

These formulas give the following asymptotics for classical plasma ($\alpha \longrightarrow -\infty$):

$$C_T^j(\alpha) = -2V_1 + \ldots, \quad C_T^Q(\alpha) = 3V_1 + \ldots,$$

$$C_U^j(p, \alpha) = 3V_1 - \frac{8p-4}{3p} + \ldots,$$

$$C_U^Q(\alpha) = -\frac{3}{2}V_1 + \ldots.$$

Figures 1–4 plot C_T^j, C_T^Q, C_U^j, and C_U^Q versus α. Coefficients C_T^Q and C_U^j are seen to reach a maximum at $\alpha \approx 4$. Coefficient C_U^j strongly depends on probability p of forward electron scattering, as follows from Fig. 3.

ACKNOWLEDGMENTS

REFERENCES

1. A. A. Abrikosov, *Introduction to the Theory of Metals* (Nauka, Moscow, 1987) [in Russian].
2. L. D. Landau and E. M. Lifshitz, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
3. A. V. Latyshev and A. A. Yushkanov, Zh. Tekh. Fiz. **73** (7), 37 (2003) [Tech. Phys. **48**, 837 (2003)].
4. A. V. Latyshev and A. A. Yushkanov, Fiz. Tverd. Tela (St. Petersburg) **43**, 1744 (2001) [Phys. Solid State **43**, 1816 (2001)].
5. L. Girifalco, *Statistical Physics of Materials* (Wiley, New York, 1973; Mir, Moscow, 1975).
6. K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, 1967; Mir, Moscow, 1972).
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1976; Pergamon, Oxford, 1980).

Translated by V. Isaakyan

THEORETICAL
AND MATHEMATICAL PHYSICS

Calculation of the Asymmetry Parameters of the Angular Distribution and of the Spin Polarization of Auger Electrons for Open-Shell Atoms

A. Yu. Elizarov* and I. I. Tupitsyn**

* Ioffe Institute for Physics and Technology, Russian Academy of Sciences,
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
e-mail: a.elizarov@mail.ioffe.ru
** St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 198904 Russia
e-mail: tup@tip.usr.pu.ru

Abstract—By means of the multichannel Hartree–Fock–Dirac method, the parameters β and α_2 of the angular distribution of Auger electrons and the parameter β_2 of their spin polarization are calculated for sodium, krypton, xenon, barium, mercury, and excited argon. Relevant Coulomb matrix elements are determined by using orthogonal multielectron initial- and final-state wave functions for intermediate-type coupling in a relativistic approximation. Exchange interaction is taken into account in all calculations. A comparison of the results of calculations performed in the frozen-core approximation and those obtained with orthogonal initial- and final-state wave functions reveals that the relaxation of core orbital processes has only a slight effect on the anisotropy parameters of the angular distribution. The resulting values of the parameters β, α_2, and β_2 are compared with the results of different calculations. © 2004 MAIK "Nauka/Interperiodica".

INTRODUCTION

Investigation of the angular distributions and spin polarization of Auger electrons furnishes information about the dynamics of Auger decay. Basic theoretical results on the subject that were based on the application of the multichannel multiconfiguration Fock–Dirac method were obtained in [1–5] for atoms containing closed shells. Calculations of the anisotropy parameters of angular distributions for atoms involving open shells are less comprehensive. In view of the aforesaid, we have calculated the parameters α_2 and β_2 for Kr and Xe (closed shells) and Na, Ba, and Hg atoms (open shells), as well as the parameter β for Ar*($2p^5 3s^2 3p^6 4s$). It is well known that calculations of the angular distributions and spin polarizations of Auger electrons are based on the two-step model of Auger decay [6]. There are a number of articles expounding on the general theory of Auger decay (see, for example, [5]). The usual expression for the angular distribution of Auger electrons can be represented in the form

$$\frac{dW_{A^+ \to A^{2+}}}{d\Omega} = \frac{dW^{\Sigma}_{A^+ \to A^{2+}}}{4\pi}[1 + \alpha_2 A_{20} P_2(\cos(\theta))], \quad (1)$$

where $dW^{\Sigma}_{A^+ \to A^{2+}}$ is the Auger process probability integrated over the directions of Auger electron trajectories, A_{20} is the population of magnetic sublevels of a singly charged ion, α_2 is the anisotropy parameter of the angular distribution of Auger electrons, P_2 is a Legendre polynomial of second degree, and θ is the angle between the direction of Auger electron emission and the polarization of the radiation.

In the case where the atom involved is excited from the state of total angular momentum $J_0 = 0$ by a linearly polarized radiation, the coefficient β can be factorized as

$$\beta = \alpha_2 A_{20}, \quad (2)$$

where $A_{20} = -\sqrt{2}$.

For the anisotropy of the angular distribution of Auger electrons and the anisotropy of their spin polarization, we have employed the expressions [4, 5]

$$\alpha_2 = \frac{A(200)}{A(000)}, \quad \beta_2 = -\frac{1}{\sqrt{3}}\frac{\Im A(211)}{A(000)}, \quad (3)$$

where

$$A(KkQ) = \frac{1}{4\pi p}\sqrt{(2K+1)(2k+1)}$$

$$\times \sum_{l,l'} i^{(l'-l)} e^{i(\sigma_l - \sigma_{l'})} \sum_{j,j'} (-1)^{J+J_1+j+Q+l'}$$

$$\times \sqrt{(2l+1)(2l'+1)(2j+1)(2j'+1)} \quad (4)$$

$$\times \begin{Bmatrix} J & J_1 & j \\ K & j' & J_1 \end{Bmatrix} \sum_X C_{l0,\,l'0}^{X0} C_{K-Q,\,kQ}^{X0} \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & \\ j' & j & K \\ l' & l & X \end{Bmatrix}$$

$$\times \langle (J, \varepsilon j) J_1 \| V \| J_1 \rangle \langle (J, \varepsilon j') J_1 \| V \| J_1 \rangle.$$

Here, J_1 is the total angular momentum of the initial state of the ion A^+; J is the total angular momentum of the final state of the ion $A^{2}+$; and j and l are, respectively, the total angular momentum and the orbital angular momentum of a partial wave of an Auger electron.

The multiconfiguration Fock–Dirac method was used in all calculations of multielectron wave functions for the singly charged ion appearing in the initial state and for the doubly charged ion appearing in the final state. The Auger electron wave function was calculated on the basis of a full relativistic method involving orthogonalization with respect to core orbitals. Exchange interaction was taken into account. A similar method of calculations was described in [7].

CALCULATION OF MATRIX ELEMENTS

In determining the transition amplitudes for Auger processes proceeding from an initial state that involves one hole in an inner shell to a final state that has two holes in inner shells, the initial- and the final-state wave function are calculated separately; therefore, they are not orthogonal to each other. This approximation is referred to as a relaxation-free approximation; in this case, core orbitals are frozen in Auger decay. In the present study, we calculate the parameters α_2 and β_2 (i) in the relaxation-free approximation and (ii) with allowance for relaxation, in which case the initial-state wave function is orthogonal to its final-state counterpart.

For an N-electron subsystem, the multielectron initial- or final-state wave function Ψ can be represented in the form of a linear combination of Slater determinants $\det_\alpha$ constructed from single-electron wave functions $\phi_i(x)$ [$\phi_f(x)$]; that is,

$$\Psi = \sum_\alpha C_\alpha \det_\alpha. \tag{5}$$

For an Auger process, where total energy E of the atom involved is conserved throughout the Auger decay process, the amplitude of the transition from the initial state $|A\rangle$ to the final state $|B\rangle$ is given by

$$\langle A | \hat{H} - e\hat{I} | B \rangle = \sum_{\alpha\beta} C_\alpha^{A*} C_\beta^B \langle \alpha | H^{AB} - ES^{AB} | \beta \rangle, \tag{6}$$

Table 1. Anisotropy parameters α_2 and β_2 of the angular distribution and spin polarization, respectively, for $L_3M_1M_{4,\,5}$ Auger transitions in Kr, Xe, Ba, and Hg ($\alpha_2^{(-)}$ and $\beta_2^{(-)}$ are the anisotropy parameters calculated in the frozen-core approximation, while $\alpha_2^{(+)}$ and $\beta_2^{(+)}$ are their counterparts calculated with allowance for relaxation in the course of Auger decay)

Element	Final-state term	α_2 [9]	$\alpha_2^{(-)}$	$\beta_2^{(-)}$	$\alpha_2^{(+)}$	$\beta_2^{(+)}$
Kr	$^1D_2$	0.218	−0.081	0.034	−0.026	0.033
	$^3D_1$	−0.034	−0.337	−0.147	−0.279	−0.156
	$^3D_2$	0.278	0.191	−0.153	0.243	−0.160
	$^3D_3$	0.331	0.612	0.147	0.570	0.141
Xe	$^1D_2$	0.228	−0.234	0.077	−0.191	0.080
	$^3D_1$	0.101	−0.422	−0.139	−0.391	−0.147
	$^3D_2$	0.342	0.584	−0.193	0.638	−0.201
	$^3D_3$	0.161	0.606	0.147	0.580	0.144
Ba	$^1D_2$	0.235	−0.211	0.076	−0.211	0.076
	$^3D_1$	0.147	−0.380	−0.152	−0.380	−0.152
	$^3D_2$	0.328	0.716	−0.186	0.716	−0.186
	$^3D_3$	0.134	0.553	0.140	0.553	0.140
Hg	$^1D_2$	0.334				
	$^3D_2$		0.025	0.073	0.393	0.073
	$^3D_1$	0.801	−0.80	−0.240	0.034	−0.250
	$^3D_2$	0.068	0.406	0.015	0.402	0.017
	$^3D_3$	−0.162	0.133	0.057	0.112	0.053

where H^{AB} is the matrix of the total Hamiltonian $\hat{H}$ in the basis of Slater determinants and S^{AB} is the orthogonality matrix.

The matrix elements of matrix S^{AB} are the matrix elements of the identity operator that is sandwiched between two Slater determinants, $\langle \det_\alpha | \det_\beta \rangle$. The matrix S^{AB} is different from the identity matrix because of the nonorthogonality of the single-electron initial- and final-state wave functions. As was shown in [8], matrix S^{AB} can be represented in the form

$$(\hat{S})_{\alpha\beta} = \langle \det_\alpha | \det_\beta \rangle = (D_{\alpha\alpha} D_{\beta\beta})^{-1/2} D_{\alpha\beta}, \tag{7}$$

where $D_{\alpha\beta} = \det|\langle \phi_i^A | \phi_j^B \rangle|$ is the determinant of the matrix of the overlap integrals $S_{i,\,j}^{\alpha\beta} = |\langle \phi_i^A | \phi_j^B \rangle|$ for two sets of orbitals $\{\phi_i^A\}_\alpha$ and $\{\phi_j^B\}_\beta$ for two Slater determinants labeled with α and β, respectively.

Further, we represent the Hamiltonian matrix for our atomic system in the form of a combination of the single- and two-particle density matrices for the transition between the states described by the Slater determi-

Table 2. Anisotropy parameter α_2 of the angular distribution for *KLL* Auger transitions in Na atoms

Term	Auger transition								
	$^3P_1$	$^3P_2$	$^1P_1$	$^3P_1$	$^3P_2$	$^1P_1$	$^3P_1$	$^3P_2$	$^1P_1$
	α_2 [10]			$\alpha_2^{(-)}$			$\alpha_2^{(+)}$		
$2s2p^{5\,2}S_{\frac{1}{2}}$	0.706	−0.837	−1.411	−0.108	−0.837	−1.411	0.707	−0.837	−1.411
$2s2p^{5\,2}P_{\frac{3}{2}}$	~0.00	0.673	0.705	0.006	0.673	0.706	−0.003	0.672	0.705
$2s2p^{5\,2}P_{\frac{1}{2}}$	−0.665	−0.837	0.705	−0.677	−0.837	0.707	−0.660	−0.837	0.705
$2s2p^{5\,2}D_{\frac{5}{2}}$	−0.141	−0.170	−0.141	−0.141	−0.170	−0.141	−0.146	−0.111	−0.142
$2s2p^{5\,2}D_{\frac{3}{2}}$	0.138	0.673	−0.141	0.144	0.673	0.141	0.139	0.672	−0.141
$2s2p^{5\,2}S_{\frac{1}{2}}$	−1.414	−0.837	−1.413	0.707	−0.837	−1.412	0.706	−0.837	−1.411
$2s2p^{5\,2}P_{\frac{1}{2}}$	−0.701	−0.837	0.701	−0.635	−0.837	0.702	−0.363	−0.837	0.705
$2s2p^{5\,2}P_{\frac{3}{2}}$	0.281	0.811	0.621	0.325	0.811	0.601	0.330	0.817	0.622
$2s2p^{5\,2}D_{\frac{3}{2}}$	−0.076	0.797	−0.061	−0.064	0.833	−0.041	−0.054	0.855	−0.019
$2s2p^{5\,2}D_{\frac{5}{2}}$	−0.141	−0.035	−0.141	0.141	−0.088	−0.141	0.139	0.672	−0.141
$2s2p^{5\,4}P_{\frac{1}{2}}$	−1.149	−0.837	−0.351	−1.115	−0.837	−0.319	−1.115	−0.837	−0.351
$2s2p^{5\,4}P_{\frac{3}{2}}$	0.652	−0.513	0.273	0.656	−0.504	0.441	0.644	−0.457	0.443
$2s2p^{5\,4}P_{\frac{5}{2}}$	−0.141	0.820	−0.141	−0.141	0.826	−0.141	−0.141	0.826	−0.142
$2s2p^{5\,4}D_{\frac{1}{2}}$	0.684	−0.837	0.698	0.683	−0.837	0.700	0.682	−0.846	0.700
$2s2p^{5\,4}D_{\frac{3}{2}}$	0.145	0.800	0.260	0.143	0.805	0.183	0.121	0.807	0.313
$2s2p^{5\,4}D_{\frac{5}{2}}$	−0.141	0.202	−0.141	−0.141	0.194	−0.141	−0.139	0.145	−0.142
$2s2p^{5\,4}D_{\frac{7}{2}}$		0.239		~0.00	−0.239	~0.00	0.707	−0.241	~0.00
$2s2p^{5\,4}S_{\frac{1}{2}}$	0.706	−0.834	0.707	0.706	−0.834	0.705	0.705	−0.833	0.706
$2s^22p^{4\,2}P_{\frac{1}{2}}$	~0.00	−0.841	~0.00	−0.690	−0.836	−0.693	−0.688	−0.836	−0.694

Table 2. (Contd.)

Term	Auger transition								
	$^3P_1$	$^3P_2$	$^1P_1$	$^3P_1$	$^3P_2$	$^1P_1$	$^3P_1$	$^3P_2$	$^1P_1$
	α_2 [10]			$\alpha_2^{(-)}$			$\alpha_2^{(+)}$		
$2s^2 2p^4\,^2P_{\frac{3}{2}}$	~0.00	~0.00	~0.00	0.511	0.010	−0.693	0.501	0.010	−0.694
$2s^2 2p^4\,^2P_{\frac{1}{2}}$	−0.702	−0.837	−0.707	−0.703	−0.837	−0.708	−0.704	−0.837	−0.708
$2s^2 2p^4\,^2P_{\frac{3}{2}}$	0.517	~0.00	−0.707	0.528	0.001	−0.708	0.518	~0.00	−0.708
$2s^2 2p^4\,^2D_{\frac{5}{2}}$	−0.144	0.589	0.705	−0.139	0.597	0.706	−0.146	0.598	0.705
$2s^2 2p^4\,^2D_{\frac{3}{2}}$	−0.573	~0.00	0.699	−0.578	~0.00	0.702	−0.571	~0.00	0.700
$2s^2 2p^4\,^2F_{\frac{7}{2}}$	−0.202	−0.240	−0.202	−0.202	−0.239	−0.202	−0.202	−0.239	−0.202
$2s^2 2p^4\,^2F_{\frac{5}{2}}$	0.196	0.598	−0.202	0.204	0.598	−0.202	0.197	0.598	−0.202
$2s^2 2p^4\,^2P_{\frac{1}{2}}$	−0.635	−0.836	−0.689	−0.690	−0.836	−0.693	−0.689	−0.836	−0.694
$2s^2 2p^4\,^2P_{\frac{3}{2}}$	0.501	0.005	−0.691	0.511	0.010	−0.693	0.501	0.010	−0.693

nants labeled with α and β. The single-particle density matrix has the form

$$\rho_1^{\alpha,\beta}(x, x') = (D_{\alpha\alpha}D_{\beta\beta})^{-1/2}$$
$$\times D_{\alpha\beta}\sum_{i,j}^{N}(S^{-1})_{i,j}^{\alpha,\beta}\phi_i^A(x)\phi_j^{B*}(x'). \tag{8}$$

The two-particle density matrix is expressed in terms of the single-particle density matrix and is given by

$$\rho_2^{\alpha,\beta}(x_1, x_2 | x_1', x_2') = (D_{\alpha\alpha}D_{\beta\beta})^{-1/2}$$
$$\times \sum_{i\neq k}^{N}\sum_{j\neq l}^{N} D_{i,j,k,l}^{\alpha\beta}\phi_i^A(x_1)\phi_j^{A*}(x_1')\phi_k^B(x_2)\phi_l^{B*}(x_2'), \tag{9}$$

where

$$D_{i,j,k,l}^{\alpha\beta} = D_{\alpha\beta}\varepsilon_{i,k}\varepsilon_{j,l}$$
$$\times [(S^{-1})_{i,j}^{\alpha,\beta}(S^{-1})_{k,l}^{\alpha,\beta} - (S^{-1})_{i,l}^{\alpha,\beta}(S^{-1})_{k,j}^{\alpha,\beta}],$$

$$\varepsilon_{i,k} = \begin{cases} 1 & i < k \\ -1 & i > k. \end{cases}$$

The Hamiltonian $\hat{H}_{A,B}$ can be represented as the sum

$$\hat{H}_{A,B} = \sum_i \hat{h}_i + \sum_{i\neq j}^{N} \hat{v}_{i,j}, \tag{10}$$

where use is made of the expressions for the single- and the two-electron components of the Hamiltonian matrix that is constructed on the basis of nonorthogonal Slater determinants [8]; that is,

$$\left\langle \alpha \left| \sum_i \hat{h}_i \right| \beta \right\rangle = (D_{\alpha\alpha}D_{\beta\beta})^{-1/2}D_{\alpha\beta}\sum_{i\neq j}^{N}(S^{-1})_{i,j}\alpha,\beta\langle i|\hat{v}|j\rangle, \tag{11}$$

$$\left\langle \alpha \left| \sum_{i\neq j} \hat{v}_{i,j} \right| \beta \right\rangle = (D_{\alpha\alpha}D_{\beta\beta})^{-1/2}$$
$$\times \sum_{i\neq k}^{N}\sum_{j\neq l}^{N} D_{i,j,k,l}^{\alpha\beta}\langle i,j|\hat{v}|k,l\rangle. \tag{12}$$

Table 3. Anisotropy parameter β of the angular distribution for $L_2, 3M_1M_{2,3}$ Auger transitions in Ar* according to the calculation in the intermediate-coupling (I) approximation with or without allowance for relaxation [$^+\beta(I)$ and $^-\beta(I)$, respectively]

$3s^33p^54s$ Term	$^+\beta(I)$	$^-\beta(I)$	β Experiment [13]	β Theory [13]	E_e, eV	E_e, eV Experiment [13]	E_e, eV Theory [13]
$^1P4s\,^2P_{\frac{3}{2}}$	−0.609	−0.701	0.11	−0.052	188.4	197.2	198.1
$^3P4s\,^2P_{\frac{3}{2}}$	−0.191	−0.195	0.09	−0.152	194.0	194.0	194.3
$^3P4s\,^4P_{\frac{3}{2}}$	0.063	0.071	−0.02	−0.032	196.4	197.7	198.7
$^1P4s\,^2P_{\frac{1}{2}}$	−0.062	−0.077	0.06	−0.003	188.4	196.1	196.5
$^3P4s\,^2P_{\frac{1}{2}}$	−0.321	−0.344	0.05	−0.031	195.6	199.3	200.2
$^3P4s\,^4P_{\frac{1}{2}}$	0.072	0.077	0.00	−0.032	196.4	199.8	200.9

RESULTS OF THE CALCULATIONS

For the $L_3M_1M_{4,5}$ Auger transitions in Ar, Xe, Ba, and Hg, the anisotropy parameters α_2 and β_2 of the angular distributions and spin polarizations, respectively, are given in Table 1. For the convenience of a comparison with experimental data, the calculations were performed for an intermediate type of coupling in the LSJ representation. In that case, the results of our calculations for the anisotropy parameters of the angular distributions and spin polarization are at odds with the results presented in [9]. The reason for this discrepancy is not quite clear. One possible reason may be associated with the calculation of the phase shifts for continuum wave functions. In Table 2, the results of the calculations for the coefficient α_2 are given for the KLL Auger transitions in Na. These results are in good agreement with those reported in [10]; nonetheless, the reasons behind the discrepancies between the values of the parameter β_2 for Auger transitions in Ar, Xe, Ba, and Hg remain unclear.

It was shown in [11] that, for atoms featuring filled shells, the relaxation of core orbitals has only a modest effect on the parameters α_2 and β_2 if the vacancies in initial-state ions are deep. Our results cited in Tables 1 and 2 lead to an analogous conclusion for atoms featuring open shells, and this justifies the use of the frozen-core approximation in calculating various parameters that characterize Auger decay.
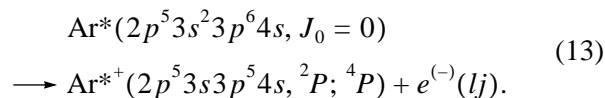
For the process

$$\text{Ar*}(2p^53s^23p^64s, J_0 = 0)$$
$$\longrightarrow \text{Ar*}^+(2p^53s3p^54s, {}^2P; {}^4P) + e^{(-)}(lj). \tag{13}$$

Table 3 gives the results of our calculations for the parameter β and for the Auger transition energies.

The results of these calculations are partly in accord with the theoretical and experimental results quoted in [13]. Many authors (see, for example, [4, 11, 12]) indicated that there are discrepancies between the theoretical and experimental results for the anisotropy parameters of the angular distributions for Auger transitions; however, a discussion of the reasons for the discrepancies between theoretical results would be premature. The numerical values of these parameters are highly sensitive to the choice of computational method; therefore, further investigations are required for revealing discrepancies between the results of different calculations.

CONCLUSIONS

The results obtained by calculating the parameters α_2 and β_2 for LMM Auger transitions in Ar*, Kr, Xe, Ba, and Hg atoms and KLL Auger transitions in Na atoms have been presented in this article. Exchange effects and the coupling between different channels of an Auger transition have been taken into account. It has been shown that the relaxation of core orbitals has but

a slight effect on the results of the calculations for the parameters α_2 and β_2.

REFERENCES

1. J. Eichler and W. Fritsch, J. Phys. B **9**, 1477 (1976).
2. E. G. Berezhko and N. M. Kabachnik, J. Phys. B **10**, 2467 (1977).
3. H. Klar, J. Phys. B **13**, 4741 (1980).
4. K. Blum, B. Lohmann, and E. Taute, J. Phys. B **19**, 3815 (1986).
5. B. Lohmann, J. Phys. B **26**, 3147 (1990).
6. W. Mehlhorn, Phys. Lett. **26**, 166 (1968).
7. A. Yu. Elizarov and I. I. Tupitsyn, Zh. Tekh. Fiz. **73** (12), 1 (2003) [Tech. Phys. **48**, 1501 (2003)].
8. R. McWeeny and B. T. Sutcliffe, *Methods of Molecular Quantum Mechanics* (Academic, New York, 1969; Mir, Moscow, 1972).
9. U. Kleiman and B. Lohmann, J. Phys. B **33**, 2653 (2000).
10. B. Lohmann, S. Fritzsche, and F. P. Larkins, J. Phys. B **29**, 1191 (1996).
11. J. Tulkki, N. M. Kabachnik, and H. Aksela, Phys. Rev. A **48**, 1277 (1993).
12. M. H. Chen, Phys. Rev. A **45**, 1684 (1992).
13. K. Ueda, Y. Shimizu, H. Chiba, *et al.*, J. Phys. B **34**, 107 (2001).

Translated by A. Isaakyan

Influence of Gas Motion Inside a Charged Bubble
in a Liquid on the Parameters of Bubble Oscillations

A. N. Zharov and A. I. Grigor'ev

Yaroslavl State University, Yaroslavl, 150000 Russia

e-mail: grig@uniyar.ac.ru

Abstract—The influence of a finite rate of leveling of the gas pressure inside a charged bubble in an ideal incompressible liquid on the bubble volume and surface oscillations is studied in a linear approximation with respect to the surface oscillation amplitude. It is shown that the bubble shape is governed by superposition of spherical harmonics with amplitudes strongly depending on their frequencies, as well as on the physical properties of the gas inside the bubble and the ambient liquid. © 2004 MAIK "Nauka/Interperiodica".

(1) Investigation of the oscillations and stability of charged vapor–gas bubbles in a dielectric liquid affected by external electric fields is of interest in connection with numerous technical and technological applications (see, for example, [1]). In this context, such bubbles have been extensively studied both experimentally [1–3] and theoretically [4–14]. Theoretical investigations of the oscillations of bubbles formed in a dielectric liquid as a result of electric [4] or laser [5] breakdown, uncharged bubbles in external electric [5−8] and magnetic [9] fields, and charged bubbles [10, 11] were mainly carried out for the radial oscillations of spherical bubbles. Recently [12], we have studied disintegration of a spheroidal charged bubble extended along the external electric field.

However, insufficient attention until now has been devoted to the surface oscillations of the bubble shape, although these very oscillations determine the mechanisms of development of the Rayleigh–Taylor instability of the surface of a collapsing bubble and the formation of a cumulative liquid jet causing cavitation erosion [2, 3, 13]. Besides, these oscillations give grounds for posing a problem concerning the translational instability of an oscillating bubble [14, 15], and they determine the mechanism of disintegration of a charged bubble [12, 16]. In [16], we studied both radial and surface oscillations of a charged bubble in a dielectric liquid in the approximation of the infinitely high sound speed in the gas inside the bubble (that is, in the approximation of the infinitely high rate of the gas pressure in the course of bubble oscillations). It should be noted that this approximation is used in the great majority of investigations of the bubble oscillations; among the papers mentioned above, only in [9] was a finite rate of leveling of the gas pressure inside a bubble taken into account. At the same time, there are many situations

where the finite sound speed in a gas may noticeably influence the laws of bubble oscillations.

The present investigation is devoted to studying volume and surface oscillations of a charged bubble with allowance for the gas motion inside it.

(2) Consider a spherical bubble with equilibrium radius r_0 formed in an ideal incompressible unbounded liquid with density $\rho^{(e)}$ and dielectric permittivity ε_d at constant pressure $P_\infty^{(e)}$ and in the absence of the gravity field. Let the bubble bear uncompensated surface charge Q and contain an ideal gas at pressure P_{0g} varying together with the volume according to the adiabatic law with an adiabatic exponent γ, and let σ denote the surface tension coefficient at the liquid–gas interface.

Under the action of the total pressure at the interface, the bubble is capable of oscillating in the vicinity of its equilibrium states so that

$$P(r) = \frac{\sigma}{r_0}\left\{\frac{r_0}{\sigma}P_{0g}\left(\frac{r_0}{r}\right)^{3\gamma} + \frac{W}{2}\left(\frac{r_0}{r}\right)^4 - \frac{2r_0}{r} - P_\infty^{(e)}\right\},$$

where

$$W \equiv \frac{Q^2}{4\pi\varepsilon_d r_0^3 \sigma}$$

and r is the current bubble radius.

When $P(r) > 0$, the bubble expands; if $P(r) < 0$, it contracts; and when $P(r) = 0$, the bubble is in equilibrium. Figure 1 presents the function $P(r)$ as presented in dimensionless variables such that $r_0 = \sigma = \rho^{(e)} = 1$. As can be seen from this figure, the equation $P(r) = 0$ may have various numbers of roots: one, two, or none [11]. In particular, a single root takes place under the condi-

tion that $\lim\limits_{r \to +\infty} P(r) = -P_\infty^{(e)} \leq 0$, that is, when the pressure of the ambient liquid is positive ($P_\infty^{(e)} \geq 0$). Two roots take place when $\lim\limits_{r \to +\infty} P(r) = -P_\infty^{(e)} > 0$ and the minimum total pressure is negative ($P_{\min} \equiv \min P(r) < 0$) and low in magnitude. To be more exact, equation $P(r) = 0$ has two roots when the pressure of the ambient liquid is slightly below zero ($P_\infty^{(e)} < 0$) and relationships $P(r_{\min}) < 0$ and $\partial_r P(r_{\min}) = 0$ are valid (symbol ∂_r denotes a derivative with respect to variable r). This state physically corresponds to a liquid in the state of undeveloped cavitation. Equation $P(r) = 0$ has no roots under the condition that $P(r_{\min}) > 0$ (provided that $P(r_{\min}) > 0$ and $\partial_r P(r_{\min}) = 0$). This state physically corresponds to developed cavitation.

Let us consider the surface oscillations of a bubble occurring in one of the equilibrium states corresponding to the condition $P(r_0) = 0$. We denote the electric field potential in the vicinity of the bubble by ϕ, and the flow velocity fields in the gas and liquid by $\mathbf{U}^{(i)}(\mathbf{r}, t)$ and $\mathbf{U}^{(e)}(\mathbf{r}, t)$, respectively. Let the gas pressure in the bubble be $P_0^{(i)}$, the gas density be $\rho_0^{(i)}$, and the sound speed be c_0. We assume that the gas inside the bubble and the ambient liquid are barotropic and obey the equations of state

$$\rho^{(i)} \equiv \rho^{(i)}(P^{(i)}) = \rho_0^{(i)}\left(\frac{P^{(i)}}{P_0^{(i)}}\right)^{1/\gamma}; \tag{1}$$

$$\rho^{(e)} = \text{const}, \tag{2}$$

where $\rho^{(i)}(\mathbf{r}, t)$ and $P^{(i)}(\mathbf{r}, t)$ are the distribution fields of the gas density and pressure inside the bubble.

We assume the bubble shape to be axisymmetric both at the initial and the subsequent time instants. The equation of the oscillating interface at any time instant t is written in the form

$$F(r, \vartheta, t) = r - r_0 - R(t) - \zeta(\vartheta, t) = 0. \tag{3}$$

The initial deformation of the bubble surface shape $[R(t) + \xi(\vartheta, t)]_{t=0}$ we assume to be such that the following relationships are fulfilled:

$$t = 0: R = \varepsilon h_0 P_0(\mu); \quad \xi = \varepsilon \sum_{m \in \Omega} h_m P_m(\mu); \tag{4}$$

$$\mu = \cos(\vartheta),$$

where ε is the amplitude of the initial perturbation ($\varepsilon \ll R$); $P_m(\mu)$ is the Legengre polynomial of the mth order; Ω is the set of indices of the initially excited modes of the surface oscillations; and h_m are dimensionless constants of the same order of smallness, which take into account the partial contribution of the mth mode to the
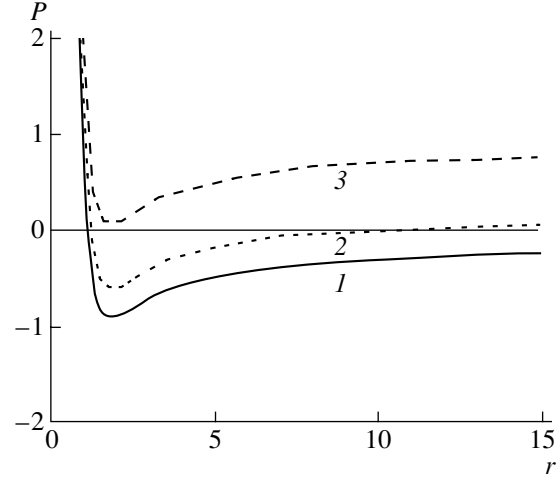
Fig. 1. Dependence of the dimensionless total pressure P at the bubble surface on the dimensionless bubble radius R for $P_{0g} = 2$, $W = 2$, $\gamma = 4/3$, and $P_\infty^{(e)} = 0.1$ (*1*), 0.2 (*2*), and 0.9 (*3*).

formation of the initial bubble surface shape such that $h_0 + \sum_{m \in \Omega} h_m = O(1)$.

The mathematical formulation of the problem of calculation of the bubble oscillations in a liquid includes, besides relations (1)–(4), the following equations and conditions:

—the Euler equations for the gas and continuity equations for the gas and liquid

$$\partial_t \mathbf{U}^{(i)} + (\mathbf{U}^{(i)} \cdot \nabla)\mathbf{U}^{(i)} = -\frac{1}{\rho^{(i)}}\nabla P^{(i)},$$

$$\partial_t \mathbf{U}^{(e)} + (\mathbf{U}^{(e)} \cdot \nabla)\mathbf{U}^{(e)} = -\frac{1}{\rho^{(e)}}\nabla P^{(e)},$$

$$\frac{d\rho^{(i)}}{dt} + \rho^{(i)}\nabla \cdot \mathbf{U}^{(i)} = 0; \quad \frac{d\rho^{(e)}}{dt} + \rho^{(e)}\nabla \cdot \mathbf{U}^{(e)} = 0;$$

—the initial conditions for the flow velocity fields

$$t = 0: \mathbf{U}^{(i)} = 0, \quad \mathbf{U}^{(e)} = 0;$$

—the Laplace equation for the electrostatic field potential

$$\Delta\phi = 0$$

(where Δ is the Laplacian);

—the boundedness conditions

$$r \longrightarrow 0: \mathbf{U}^{(i)} \longrightarrow 0,$$

$$r \longrightarrow +\infty: \mathbf{U}^{(e)} \longrightarrow 0, \quad \nabla\phi \longrightarrow 0;$$

—the kinematic and dynamic boundary conditions

$$r = r_0 + R(t) + \xi(\vartheta, t): \frac{dF}{dt} = 0,$$

$$P^{(i)} + P_q = P^{(e)} + P_\sigma;$$

—the condition of invariability of the total charge

$$\int\limits_S \mathbf{n} \cdot \boldsymbol{\nabla}\phi \, dS = -4\pi Q,$$

$$S = \{r, \vartheta, \varphi \,|\, r = r_0 + R(t) + \xi(\vartheta; t);$$
$$0 \le \vartheta \le \pi; \, 0 \le \varphi \le 2\pi\};$$

—and the condition of constancy of the electric potential at the bubble surface

$$r = r_0 + R(t) + \xi(\vartheta, t): \phi = \phi_S(t).$$

In the expressions written above, P_q and P_σ are the electric field pressure and capillary pressure, respectively; $\mathbf{n}$ is the unit vector normal to the bubble surface; ϕ_S is the electric potential of the bubble surface; and ∂_t denotes a partial derivative with respect to variable t.

(3) In the following, the gas and liquid flows are assumed to be potential, with the corresponding velocity potentials $\psi^{(i)}$ and $\psi^{(e)}$. Substituting $\mathbf{U}^{(e)} = \boldsymbol{\nabla}\psi^{(e)}$ into the Euler equation and the continuity equation for the ambient medium and into the initial condition, and taking into account the equation of state of liquid (2), we obtain the pressure distribution in the liquid and the Laplace equation for the velocity field potential $\psi^{(e)}$,

$$P^{(e)} = P_\infty^{(e)} - \rho^{(e)}\left(\partial_t\psi^{(e)} + \frac{1}{2}(\boldsymbol{\nabla}\psi^{(e)})^2\right); \quad \Delta\psi^{(e)} = 0.$$

Substituting $\mathbf{U}^{(i)} = \boldsymbol{\nabla}\psi^{(i)}$ into the Euler equation for the gas flow inside the bubble and integrating over the bubble volume, we obtain the Cauchy integral

$$\partial_t\psi^{(i)} + \frac{1}{2}(\boldsymbol{\nabla}\psi^{(i)})^2 + \Pi^{(i)} = 0, \tag{5}$$

where $\Pi^{(i)}$ is a function satisfying the relation

$$d\Pi^{(i)} = \frac{1}{\rho^{(i)}}dP^{(i)}. \tag{6}$$

Substituting expression $\mathbf{U}^{(i)} = \boldsymbol{\nabla}\psi^{(i)}$ into the continuity equation for the gas flow and taking into account (6) and relation $c^2 = dP^{(i)}/d\rho^{(i)}$, where c is the sound speed, we obtain the equation

$$\frac{1}{c^2}\frac{d\Pi^{(i)}}{dt} + \Delta\psi^{(i)} = 0. \tag{7}$$

Expressing function $\Pi^{(i)}$ from (5) and substituting it into (7), we obtain the nonlinear equation for finding

the gas velocity field potential

$$\Delta\psi^{(i)} - \frac{1}{c^2}\partial_{t,t}\psi^{(i)} - \frac{2}{c^2}(\boldsymbol{\nabla}\psi^{(i)})(\boldsymbol{\nabla}\partial_t\psi^{(i)})$$
$$- \frac{1}{2c^2}(\boldsymbol{\nabla}\psi^{(i)})(\boldsymbol{\nabla}(\boldsymbol{\nabla}\psi^{(i)})^2) = 0 \tag{8}$$

with the initial condition

$$t = 0: \psi^{(i)} = 0. \tag{9}$$

Note that nonlinear wave equation (8) contains partial derivatives of the second order of the gas velocity potential with respect to time, so that single initial condition (9) is insufficient for finding the solution. Therefore, Eq. (8) will be supplemented by the physically evident condition

$$\partial_t\psi^{(i)} = 0. \tag{10}$$

Integrating (6) over the bubble volume subject to the equation of state (1) of the gas, we find the expression for $\Pi^{(i)}$

$$\Pi^{(i)} = \frac{\gamma}{\gamma - 1}\left(\frac{P^{(i)}}{\rho^{(i)}} - \frac{P_0^{(i)}}{\rho_0^{(i)}}\right). \tag{11}$$

Substituting (11) into (5), taking into account Eq. (1), and using the relation $c_0^2 = \gamma P_0^{(i)}/\rho_0^{(i)}$, we find the gas pressure distribution and sound speed in the bubble

$$P^{(i)} = P_0^{(i)}\left(1 - \frac{\gamma - 1}{c_0^2}\left(\partial_t\psi^{(i)} + \frac{1}{2}(\boldsymbol{\nabla}\psi^{(i)})^2\right)\right)^{\gamma/\gamma - 1};$$

$$c^2 = c_0^2 - (\gamma - 1)\left(\partial_t\psi^{(i)} + \frac{1}{2}(\boldsymbol{\nabla}\psi^{(i)})^2\right).$$

The solution of the nonlinear set of equations considered is sought by linearization in the form of expansions in small parameter ε

$$R(t) = \varepsilon R^{(1)}(t) + O(\varepsilon^2);$$
$$\xi(\vartheta, t) = \varepsilon\xi^{(1)}(\vartheta, t) + O(\varepsilon^2);$$
$$\phi(r, \vartheta, t) = \varepsilon\phi(r, \vartheta, t) + O(\varepsilon^2); \tag{12}$$
$$\psi^{(i)}(r, \vartheta, t) = \varepsilon\psi^{(i)(1)}(r, \vartheta, t) + O(\varepsilon^2);$$
$$\psi^{(e)}(r, \vartheta, t) = \varepsilon\psi^{(e)(1)}(r, \vartheta, t) + O(\varepsilon^2).$$

(4) Having substituted expansions (12) into the problem formulated above, let us separate a problem of the zeroth order of smallness in ε:

$$\Delta\phi^{(0)} = 0; \quad r \longrightarrow +\infty: \boldsymbol{\nabla}\phi^{(0)} \longrightarrow 0;$$
$$r = r_0: P^{(i)} + P_q = P^{(e)} + P_\sigma;$$

$$\int_{-1}^{1} r_0^2 \partial_r \phi^{(0)} d(\cos\vartheta) = -2Q; \quad \phi^{(0)} = \phi_S^{(0)}(t).$$

Solving this set, we find

$$\phi^{(0)} = \frac{Q}{r}; \quad \phi_S^{(0)} = \frac{Q}{r_0}; \quad P_0^{(i)} + \frac{Q^2}{8\pi\varepsilon_d r_0^4} = \frac{2\sigma}{r_0} + P_\infty^{(e)}. \quad (13)$$

Relations (13) determine the magnitude of equilibrium bubble radius r_0.

(5) Separating the terms including the small parameter in the first power, we obtain a problem of the first order of smallness in the form

$$\Delta\phi^{(1)} = 0; \quad \Delta\psi^{(e)(1)} = 0; \quad \Delta\psi^{(i)(1)} - \frac{1}{c_0^2}\partial_{t,t}\psi^{(i)(1)} = 0;$$

$$t = 0: \psi^{(i)(1)} = 0; \quad \partial_t\psi^{(i)(1)} = 0;$$

$$R^{(1)} = h_0 P_0(\mu); \quad \xi^{(1)} = \sum_{m \in \Omega} h_m P_m(\mu);$$

$$r \longrightarrow 0: \psi^{(i)(1)} \longrightarrow 0;$$

$$r \longrightarrow +\infty: \psi^{(e)(1)} \longrightarrow 0; \quad \nabla\phi^{(1)} \longrightarrow 0;$$

$$r = r_0: \phi^{(1)} + (R^{(1)} + \xi^{(1)})\partial_r\phi^{(0)} = \phi_S^{(1)}(t);$$

$$\partial_r\psi^{(i)(1)} = \partial_r\psi^{(e)(1)} = \partial_t R^{(1)} + \partial_t\xi^{(1)};$$

$$-\rho_0^{(i)}\partial_t\psi^{(i)(1)} + \frac{1}{4\pi\varepsilon_d}\partial_r\phi^{(0)}(\partial_r\phi^{(1)} + (R^{(1)} + \xi^{(1)})\partial_{rr}\phi^{(0)})$$

$$= -\rho^{(e)}\partial_t\psi^{(e)(1)} - \frac{2\sigma}{r_0^2}R^{(1)} - \frac{\sigma}{r_0^2}(2 + \Delta_\Omega)\xi^{(1)};$$

$$\int_{-1}^{1}(r_0\partial_r\phi^{(1)} + (R^{(1)} + \xi^{(1)}) \\ \times (r_0\partial_{rr}\phi^{(0)} + 2\partial_r\phi^{(0)}))d(\cos\vartheta) = 0. \quad (14)$$

Since set of equations (14) is linear, we will find its solution using the Fourier transform F_+ with respect to time, that is, by passing from functions $\phi^{(1)}(r, \vartheta, t)$, $\psi^{(e)(1)}(r, \vartheta, t)$, $\psi^{(i)(1)}(r, \vartheta, t)$, $R^1(t)$, and $\xi^{(1)}(\vartheta, t)$ to their Fourier transforms [17]

$$F_+(f) = \int_0^{+\infty} f e^{iSt} dt, \quad f = \phi^{(1)}, \quad f = \psi^{(e)(1)},$$

$$f = \psi^{(i)(1)}, \quad f = R^{(1)}, \quad f = \xi^{(1)}.$$

Then, set (14) takes the form

$$\Delta F_+(\phi^{(1)}) = 0; \quad \Delta F_+(\psi^{(e)(1)}) = 0; \quad (15)$$

$$\Delta F_+(\psi^{(i)(1)}) + \frac{S^2}{c_0^2}F_+(\psi^{(i)(1)}) = 0; \quad (16)$$

$$r \longrightarrow 0: F_+(\psi^{(i)(1)}) \longrightarrow 0; \quad (17)$$

$$r \longrightarrow +\infty: F_+(\psi^{(e)(1)}) \longrightarrow 0; \quad \nabla F_+(\phi^{(1)}) \longrightarrow 0; (18)$$

$$r = r_0: \partial_r F_+(\psi^{(i)(1)}) = \partial_r F_+(\psi^{(e)(1)}) = -iSF_+(R^{(1)})$$

$$-h_0 P_0(\mu) - iSF_+(\xi^{(1)}) - \sum_{m \in \Omega} h_m P_m(\mu); \quad (19)$$

$$\rho_0^{(i)}iSF_+(\psi^{(i)(1)}) + \frac{1}{4\pi\varepsilon_d}\partial_r\phi^{(0)}(\partial_r F_+(\phi^{(1)})$$

$$+ (F_+(R^{(1)}) + F_+(\xi^{(1)}))\partial_{rr}\phi^{(0)}) \quad (20)$$

$$= \rho^{(e)}iSF_+(\psi^{(e)(1)}) - \frac{2\sigma}{r_0^2}F_+(R^{(1)}) - \frac{\sigma}{r_0^2}(2 + \Delta_\Omega)F_+(\xi^{(1)});$$

$$\int_{-1}^{1}(r_0\partial_r F_+(\phi^{(1)}) + (F_+(R^{(1)}) + F_+(\xi^{(1)}))$$

$$\times (r_0\partial_{rr}\phi^{(0)} + 2\partial_r\phi^{(0)}))d(\cos\vartheta) = 0; \quad (21)$$

$$F_+(\phi^{(1)}) + (F_+(R^{(1)}) + F_+(\xi^{(1)}))\partial_r\phi^{(0)} = F_+(\phi_s^{(1)}). \quad (22)$$

To find the solution of set (15)–(22), we expand the Fourier transform of the deviation of the bubble surface from the equilibrium shape in terms of the Legendre polynomials

$$F_+(R^{(1)}) + F_+(\xi^{(1)}) = A_0^{(1)}(S) + \sum_{n=1}^{\infty} A_n^{(1)}(S)P_n(\mu).$$

Then, the solution of Eqs. (15) and (16) subject to boundedness conditions (17), (18) and conditions (21), (22) can be written in the following form:

$$F_+(\phi^{(1)}) = \frac{Q}{r_0^2}\sum_{n=1}^{\infty} A_n^{(1)}(S)\left(\frac{r_0}{r}\right)^{n+1}P_n(\mu);$$

$$F_+(\phi_S^{(1)}) = -\frac{Q}{r_0^2}A_0^{(1)}(S);$$

$$F_+(\psi^{(e)(1)}) = \sum_{n=0}^{\infty} C_n^{(e)(1)}(S)\frac{1}{r^{n+1}}P_n(\mu);$$

$$F_+(\psi^{(i)(1)}) = \sum_{n=0}^{\infty} C_n^{(i)(1)}(S)j_n\left(\frac{S}{c_0}r\right)P_n(\mu), \quad (23)$$

where $C_n^{(i)(1)}$, $C_n^{(e)(1)}$ are constants and $j_n(Sr/c_0)$ are spherical Bessel functions of the first kind.

Substituting expressions (23) into Eqs. (19) and (20), and taking into account the mutual orthogonality of the Legendre polynomials, we obtain

$$
\begin{cases}
C_0^{(i)(1)}\partial_{r_0}j_0\left(\dfrac{S}{c_0}r_0\right) = -C_0^{(e)(1)}\dfrac{1}{r_0^2} = iSA_0^{(1)} - h_0, \\[2mm]
\rho_0^{(i)}iSC_0^{(i)(1)}j_0\left(\dfrac{S}{c_0}r_0\right) - \dfrac{\rho^{(e)}}{r_0}iSC_0^{(e)(1)} + \dfrac{2\sigma}{r_0^2}(1-W)A_0^{(1)} = 0;
\end{cases} \tag{24}
$$

$$
\begin{cases}
C_n^{(i)(1)}\partial_{r_0}j_n\left(\dfrac{S}{c_0}r_0\right) = -C_n^{(e)(1)} + \dfrac{n+1}{r_0^{n+2}} = -iSA_n^{(1)} - h_n, \\[2mm]
\rho_0^{(i)}iSC_n^{(i)(1)}j_n\left(\dfrac{S}{c_0}r_0\right) - \dfrac{\rho^{(e)}}{r_0^{n+1}}iSC_n^{(e)(1)} \\[2mm]
\quad - \dfrac{\sigma}{r_0^2}(n-1)((n+2)-W)A_n^{(1)} = 0.
\end{cases} \tag{25}
$$

Set of equations (24) describes radial oscillations of the bubble surface, and set (25) describes the lateral oscillations.

From set (24) subject to the relation $\partial_\chi j_n(\chi) = nj_n(\chi)/\chi - j_{n+1}(\chi)$, we find expressions for $C_0^{(i)(1)}$, $C_0^{(e)(1)}$, and $A_0^{(1)}$:

$$
A_0^{(1)}(s) = \frac{iSh_0}{D_0(S)}\left(\frac{\rho_0^{(i)}c_0}{\rho^{(e)}r_0 S}j_0(\chi) - j_1(\chi)\right); \tag{26}
$$

$$
C_0^{(i)(1)}(S) = -\frac{c_0\alpha_0 h_0}{SD_0(S)}; \quad C_0^{(e)(1)}(S) = -\frac{r_0^2\alpha_0 h_0}{D_0(S)}j_1\chi;
$$
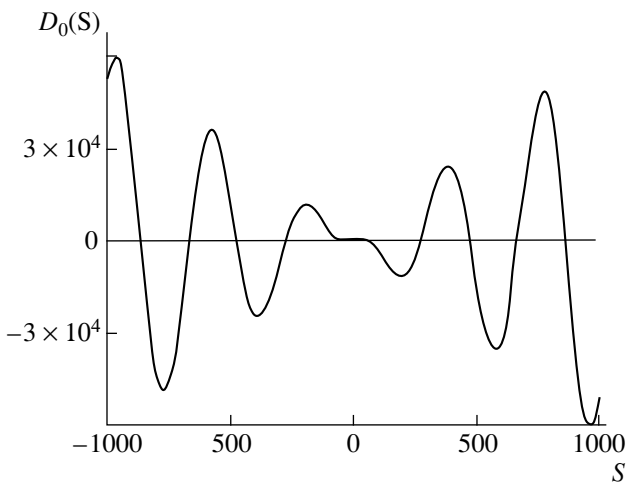
Fig. 2. Dependence of the dimensionless function $D_0(S)$ on dimensionless frequency S for $c_0 = 50$, $W = 1$, $P_{0g} = 1$, $P_\infty^{(e)} = 1$, $\gamma = 4/3$.

$$
D_0(S) = \frac{c_0\rho_0^{(i)}}{r_0\rho^{(e)}}Sj_0(\chi) - (S^2 + \alpha_0)j_1(\chi); \tag{27}
$$

$$
\chi = \frac{Sr_0}{c_0}; \quad \alpha_0 = \frac{2\sigma}{\rho^{(e)}r_0^3}(1-W).
$$

From set (25), we find $C_n^{(i)(1)}$, $C_n^{(e)(1)}$, and $A_n^{(1)}$ for any $n \geq 2$

$$
A_n^{(1)}(S)
$$
$$
= \frac{iSh_n}{D_n(S)}\left(\frac{c_0}{Sr_0}\left(\frac{r_0^{(i)}}{\rho^{(e)}}(n+1)+n\right)j_n(\chi) - j_{n+1}(\chi)\right); \tag{28}
$$

$$
C_n^{(i)(1)}(S) = \frac{c_0}{SD_n(S)}\omega_n^2 h_n;
$$

$$
C_n^{(e)(1)}(S) = -\frac{r_0^{n+2}\omega_n^2 h_n}{(n+1)D_n(S)}\left(\frac{nc_0}{Sr_0}j_n(\chi) - j_{n+1}(\chi)\right);
$$

$$
D_n(S) = \frac{c_0}{r_0}\left(\frac{\rho_0^{(i)}}{\rho^{(e)}}(n+1)S + n\left(S - \frac{\omega_n^2}{S}\right)\right)j_n(\chi) \tag{29}
$$
$$
\quad - (S^2 - \omega_n^2)j_{n+1}(\chi);
$$

$$
\omega_n^2 = \frac{\sigma}{\rho^{(e)}r_0^3}(n-1)(n+1)((n+2)-W).
$$

From expressions (26)–(29) and Fig. 2, showing the plot of dependence $D_n(S)$, it can be seen that they have the infinite but countable number of singular points $S = \pm S_{n,k}$, $n \in (\{0\} \cup \Omega)$, $k = 0, \pm 1, \pm 2, \ldots$, satisfying condition $D_n(S) = 0$ and located on either real or imaginary axes. Subscript n numbers the modes of the surface oscillations, and subscript k the modes of the radial bubble oscillations corresponding to distinct roots of the equation $D_n(S) = 0$. In expressions (26), (27), singular points $S = \pm S_{n,k}$ are simple poles, except for point $S = 0$, which is a pole of the second order. Besides, all the Fourier transforms tend to zero at $S \longrightarrow \infty$. In connection with the above considerations, the inverse Fourier transform has the following form [17]:

$$
f(t) = \frac{1}{2\pi}\int_{-\infty+i\tau}^{+\infty+i\tau} F_+(S)e^{-iSt}dS \quad (t > 0),
$$

where τ is a positive real number. This formula can be transformed into an integral along a contour enclosing the whole complex plane below the straight line $\text{Im}S = \tau$. Applying the theorem of residues to this contour integral, we can rewrite the formula of inversion as

$$
f(t) = -i\sum_{k=-\infty}^{+\infty}\underset{S=S_k}{\text{Re}}SF_+(S)e^{-iSt}. \tag{30}
$$

Using relation (30) and recurrent formulas for the spherical Bessel functions $\partial_\chi j_n(\chi) = n j_n(\chi)/\chi - j_{n+1}(\chi)$ and $\partial_\chi j_n(\chi) = j_{n-1}(\chi) - (n+1)j_n(\chi)/\chi$, we find, from (26) and (28),

$$R^{(1)}(t) = h_0 A_0(0) + \sum_{k=1}^{+\infty} h_0 A_0(S_{0,k}) \cos(S_{0,k}t); \quad (31)$$

$$\xi^{(1)}(t, \vartheta) = \sum_{n \in \Omega} \sum_{k=1}^{+\infty} h_n A_n(S_{n,k}) \cos(S_{n,k}t) P_n(\mu);$$

$$A_0(0) = 3\frac{\rho_0^{(i)}}{\rho^{(e)}}\left(\frac{c_0}{r_0}\right)^2 \left[3\frac{\rho_0^{(i)}}{\rho^{(e)}}\left(\frac{c_0}{r_0}\right)^2 - \alpha_0\right]^{-1};$$

$$A_0(S) = 2\frac{S}{\partial_S D_0(S)}\left(\frac{c_0}{Sr_0}\frac{\rho_0^{(i)}}{\rho^{(e)}}j_0(\chi) - j_1(\chi)\right);$$

$$A_n(S) = 2\frac{S}{\partial_S D_n(S)}$$

$$\times\left(\frac{c_0}{Sr_0}\left(\frac{\rho_0^{(i)}}{\rho^{(e)}}(n+1) + n\right)j_n(\chi) - j_{n+1}(\chi)\right); \quad (32)$$

$$\partial_S D_0(S) = \left(\frac{c_0\rho_0^{(i)}}{r_0\rho^{(e)}} - \frac{r_0}{c_0}(S^2 + \alpha_0)\right)j_0(\chi)$$

$$-\left(\frac{\rho_0^{(i)}}{\rho^{(e)}}S - 2\frac{\alpha_0}{S}\right)j_1(\chi);$$

$$\partial_S D_n(S) = \left(\frac{c_0}{r_0}\left(\frac{\rho_0^{(i)}}{\rho^{(e)}}(n+1)^2 + n(n+1) - n(n-1)\frac{\omega_n^2}{S^2}\right)\right.$$

$$\left.- \frac{r_0}{c_0}(S^2 - \omega_n^2)\right)j_n(\chi) - \left(\frac{\rho_0^{(i)}}{\rho^{(e)}}(n+1)S + 2\frac{\omega_n^2}{S}\right)j_{n+1}(\chi),$$

where $S_{0,k}$ is the root of the equation $D_0(S) = 0$ and $S_{n,k}$ is the root of the equation $D_n(S) = 0$.

Substituting expressions (31) and (32) into Eq. (3), we obtain an equation of the bubble generatrix:

$$r = r_0 + \varepsilon h_0 M_0(t) + \varepsilon \sum_{n \in \Omega} h_n M_n(t) P_n(\mu),$$

$$M_0(t) = A_0(0) + \sum_{k=1}^{+\infty} A_0(S_{0,k}) \cos(S_{0,k}t), \quad (33)$$

$$M_n(t) = \sum_{k=1}^{+\infty} A_n(S_{n,k}) \cos(S_{n,k}t).$$

From Eq. (33), it is seen that, inasmuch as in the general case $A_0(S_{0,k}) \neq 1$ and $A_n(S_{n,k}) \neq 1$, finiteness of the sound speed in the gas influences both the amplitudes of the radial and surface modes of oscillations and the shape of the oscillating bubble.

Substituting constants $C_0^{(i)(1)}$, $C_0^{(e)(1)}$, $C_n^{(i)(1)}$, and $C_n^{(e)(1)}$ defined by expressions (27) and (29) into expressions (23) and employing formula of inversion (30), we find expressions for the potentials of the gas and liquid fields:

$$\psi^{(i)(1)}(r, \vartheta, t) = \varepsilon C_0^{(i)}(0) h_0 t$$

$$+ \varepsilon \sum_{k=1}^{+\infty} h_0 C_0^{(i)}(S_{0,k}) j_0\left(\frac{S_{0,k}}{c_0}r\right)\sin(S_{0,k}t)$$

$$+ \varepsilon \sum_{n \in \Omega} \sum_{k=1}^{+\infty} h_n C_n^{(i)}(S_{n,k}) j_n\left(\frac{S_{n,k}}{c_0}r\right)\sin(S_{n,k}t) P_n(\mu);$$

$$\psi^{(e)(1)}(r, \vartheta, t) = \varepsilon \sum_{k=1}^{+\infty} h_0 \frac{C_0^{(e)}(S_{0,k})}{r}\sin(S_{0,k}t)$$

$$+ \varepsilon \sum_{n \in \Omega} \sum_{k=1}^{+\infty} h_n \frac{C_n^{(e)}(S_{n,k})}{r^{n+1}}\sin(S_{n,k}t) P_n(\mu);$$

$$(34)$$

$$C_0^{(i)}(0) = 3\frac{c_0^2}{r_0}\alpha_0\left[3\frac{\rho_0^{(i)}}{\rho^{(e)}}\left(\frac{c_0}{r_0}\right)^2 - \alpha_0\right]^{-1};$$

$$C_0^{(i)}(S) = 2\frac{c_0}{S\partial_S D_0(S)}\alpha_0;$$

$$C_n^{(i)}(S) = -2\frac{c_0}{S\partial_S D_n(S)}\omega_n^2;$$

$$C_0^{(e)}(S) = 2r_0^2\alpha_0\frac{j_1(\chi)}{\partial_S D_0(S)};$$

$$C_n^{(e)}(S) = 2\frac{r_0^{n+2}}{(n+1)}\frac{\omega_n^2}{\partial_S D_n(S)}\left(\frac{nc_0}{Sr_0}j_n(\chi) - j_{n+1}(\chi)\right).$$

(6) Inasmuch as a bubble in a liquid obeys the relation $Sr_0/c_0 = \chi \ll 1$, this situation is worthy of separate consideration. In this case, the following relations are valid for the spherical Bessel functions:

$$j_0(\chi) = 1 - \frac{\chi^2}{6} + O(\chi^3); \quad j_1(\chi) = \frac{\chi}{3} + O(\chi^3);$$

$$(35)$$

$$j_n(\chi) = \frac{\chi^n}{(2n+1)!!} + O(\chi^{n+2}); \quad j_{n+1}(\chi) = O(\chi^{n+1}).$$

Substituting relations (35) into the expressions for $D_0(S)$ and $D_n(S)$, we can rewrite these expressions as

$$D_0(S) = S(S_{0,1}^2 - S^2)\left(2 + \frac{\rho_0^{(i)}}{\rho^{(e)}}\right)\frac{r_0}{6c_0} + O(\chi^3);$$

$$D_n(S) = \frac{1}{(2n+1)!!}\left(\frac{S}{c_0}r_0\right)^{n-1}\left(\frac{\rho_0^{(i)}}{\rho^{(e)}}(n+1) + n\right) \quad (36)$$

$$\times (S^2 - S_{n,1}^2) + O(\chi^{n+1}).$$

It is seen that each of Eqs. (36) has a single root for the frequency squared corresponding to $k = 1$:

$$S_{0,1}^2 = \frac{6\rho_0^{(i)}c_0^2 - 2\rho^{(e)}r_0^2\alpha_0}{(2\rho^{(e)} + \rho_0^{(i)})r_0^2};$$

$$S_{n,1}^2 = \frac{n\rho^{(e)}\omega_n^2}{\rho_0^{(i)}(n+1) + \rho^{(e)}n}. \quad (37)$$

Note that, in formula (37), the expression for $S_{0,1}^2$ differs only by the denominator from the square of the frequency of radial oscillations of the bubble calculated with neglect of the inertial gas properties in the bubble.

In particular, substituting formally $2\rho^{(e)}$ for $(2\rho^{(e)} + \rho_0^{(i)})$ in the denominator of the expression for $S_{0,1}^2$ and taking into account that $c_0^2 = \gamma P_0^{(i)}/\rho_0^{(i)}$, we obtain the well-known expression for the square of frequency of the radial oscillations of a charged bubble in the absence of gas motions inside it [10, 18]:

$$\omega_0^2 = \frac{3\gamma P_0^{(i)}}{\rho^{(e)}r_0^2} - \frac{2\sigma}{\rho^{(e)}r_0^3}(1 - W).$$

It can be also seen that the expression for $S_{n,1}^2$ exactly coincides with that for the frequency of capillary oscillations of a charged drop immersed in an ambient medium [10, 18].

Substituting expressions (35), (36) into (26)–(29) and taking into account relations (23), (30), and (37), we readily find expressions for the bubble surface shape and the potentials of the gas and liquid velocity fields in the asymptotic situation considered:

$$r = r_0 + \varepsilon h_0 \left(3\frac{\rho_0^{(e)}}{\rho^{(e)}}\left(\frac{c_0}{r_0}\right)^2 - \alpha_0\right)^{-1}$$

$$\times\left(3\frac{\rho_0^{(e)}}{\rho^{(e)}}\left(\frac{c_0}{r_0}\right)^2 - \alpha_0\cos(S_{0,1}t)\right) + \varepsilon\sum_{n\in\Omega}h_n\cos(S_{n,1}t)P_n(\mu);$$

$$\psi^{(i)(1)}(r,\vartheta,t) = \varepsilon\frac{6\rho^{(e)}\alpha_0c_0^2h_0}{(2\rho^{(e)} + \rho_0^{(i)})r_0S_{0,1}^2}$$

$$\times\left(t - \frac{\sin(S_{0,1}t)}{S_{0,1}} + \frac{S_{0,1}r^2}{6c_0^2}\sin(S_{0,1}t)\right) \quad (38)$$

$$-\varepsilon\sum_{n\in\Omega}\frac{1}{n}\left(\frac{r}{r_0}\right)^n r_0 S_{n,1}h_n\sin(S_{n,1}t)P_n(\mu);$$

$$\psi^{(e)(1)} = -\varepsilon\frac{2r_0\alpha_0\rho^{(e)}h_0}{(2\rho^{(e)} + \rho_0^{(i)})S_{0,1}}\frac{r_0}{r}\sin(S_{0,1}t)$$

$$+\varepsilon\sum_{n\in\Omega}\frac{1}{n+1}\left(\frac{r_0}{r}\right)^{n+1} r_0 S_{n,1}\sin(S_{n,1}t)P_n(\mu).$$

If $h_0 = 0$ in expressions (38), we obtain the well-known expressions describing the surface oscillations of a charged drop in an ambient medium [10, 18]. When $h_0 \neq 0$, expressions (38) approximately describe bubble oscillations at the basic (lowest) frequency.

(7) For convenience of numerical analysis of the solution obtained, let us pass to dimensionless variables by setting $\rho^{(e)} = \sigma = r_0 = 1$. Then all physical quantities of the problem are expressed in terms of their characteristic scales. Namely, for the scales of length, density, time, frequency, and velocity we respectively have

$$r_0, \rho^{(e)}, \sqrt{\frac{\rho^{(e)}r_0^3}{\sigma}}, \sqrt{\frac{\sigma}{\rho^{(e)}r_0^3}}, \sqrt{\frac{\sigma}{\rho^{(e)}r_0}}, \frac{\sigma}{r_0}.$$

According to the experimental data, the radius of a bubble in a liquid ranges from $r_0 = 10^{-1}$ to 10^{-7} cm. The surface tensions and densities of many liquids are slightly different from $\sigma = 50$ dyn/cm and $\rho^{(e)} = 1$ g/cm^3, respectively. For such values of the physical quantities, the characteristic time scale varies in the range from 5×10^{-12} to 5×10^{-12} s; the frequency scale, from 2×10^2 to 2×10^{11} s^{-1}; the velocity scale, from 2×10^1 to 2×10^4 cm/s; and the pressure scale, from 5×10^2 to 5×10^8 dyn/cm^2.

The free parameters of the problem are P_{0g}, $P_\infty^{(e)}$, W, c_0, γ, ε, Ω, and h_n, where $n \in (\Omega \cup \{0\})$. The equilibrium gas pressure in the bubble is defined by expression $P_0^{(i)} = P_{0g}r_0^{-3\gamma}$, and the gas density is $\rho_0^{(i)} = \gamma P_0^{(i)}/c_0^2$.

In the course of numerical analysis of the abovesolution describing oscillations of a charged bubble in a liquid, it was established that equations $D_n(S) = 0$, where $n \in (\Omega \cup \{0\})$, have an infinite number of solutions $S = \pm S_{n,k}$, $k = 0, \pm 1, \pm 2, \ldots$ which define the frequency (the growth rate) of the bubble surface oscillations.

For example, when a bubble in a liquid has a single equilibrium state (i.e., Eq. (13) has a single solution), all solutions of the equation $D_0(S) = 0$ determining the radial bubble oscillations are real and determine the frequencies of radial bubble oscillations, which are strongly dependent on sound speed c_0 (Fig. 3), liquid pressure $P_\infty^{(e)}$, gas pressure P_{0g} in the bubble, and parameter W. Increasing P_{0g} and W results in decreasing the frequencies, while increasing $P_\infty^{(e)}$ and c_0 causes an increase in the frequencies. Note that the relations $S_{0,k} = S_{0,k}(c_0)$, where $S_{0,k}$ is the frequency of the kth harmonic of the radial oscillations, have the form of saturation curves. For example, at $P_\infty^{(e)} = 1$, $P_{0g} = 0.9$, $W = 1$, and $\gamma = 4/3$, the lowest harmonic of the radial oscillations ($k = 1$) ceases to depend on c_0 already at $c_0 > 10$ and represents a straight line, and the sum of the amplitudes of deviations of the bubble radius from the equilibrium magnitude $A_0(0) + A_0(S_{0,1})$ becomes equal to unity, which corresponds to the limiting case of the incompressible gas. Such a form of the dependence $S_{0,k} = S_{0,k}(c_0)$ is explained by the fact that, as sound speed c_0 increases, gas density $\rho_0^{(i)} = \gamma P_0^{(i)}/c_0^2$ in the bubble rises. Note also that, at any physically admissible values of the parameters of the problem, the amplitudes of deviations of the bubble radius from the equilibrium magnitude $A_0(S)$ vary in such a way that the sum $A_0(0) + A_0(S_{0,1})$ is slightly different from unity, and the amplitudes of higher harmonics of the radial oscillations always remain at least by an order of magnitude smaller than unity and strongly depend on sound speed c_0 (the amplitudes decrease when this speed increases, see Fig. 3b).

The equations $D_n(S) = 0$, determining the surface bubble oscillations, have not only real solutions but also imaginary ones in the above-mentioned ranges of the physical parameters. The real solutions determine the frequencies of the surface oscillations, which (being the radial oscillations) depend on the sound speed c_0, the liquid pressure $P_\infty^{(e)}$, the gas pressure P_{0g} in the bubble, and the parameter W. As the sound speed c_0 increases, the frequencies of the surface oscillations grow and tend to their limiting values corresponding to the limiting case of the incompressible gas. When this takes place, as in the case of radial oscillations, the lowest harmonic ($k = 1$) of oscillations first reaches its limiting magnitude. For example, at $n = 2$, $P_\infty^{(e)} = 1$, $P_{0g} = 0.4$, $W = 1$, and $\gamma = 4/3$, the lowest harmonic $S_{2,1}$ of the surface oscillations ceases to depend on c_0 even at $c_0 \geq 10$ and amplitude $A_2(S_{2,1})$ of deviation of the bubble surface from spherical shape for this harmonic asymptotically tends to unity. Harmonics with higher numbers
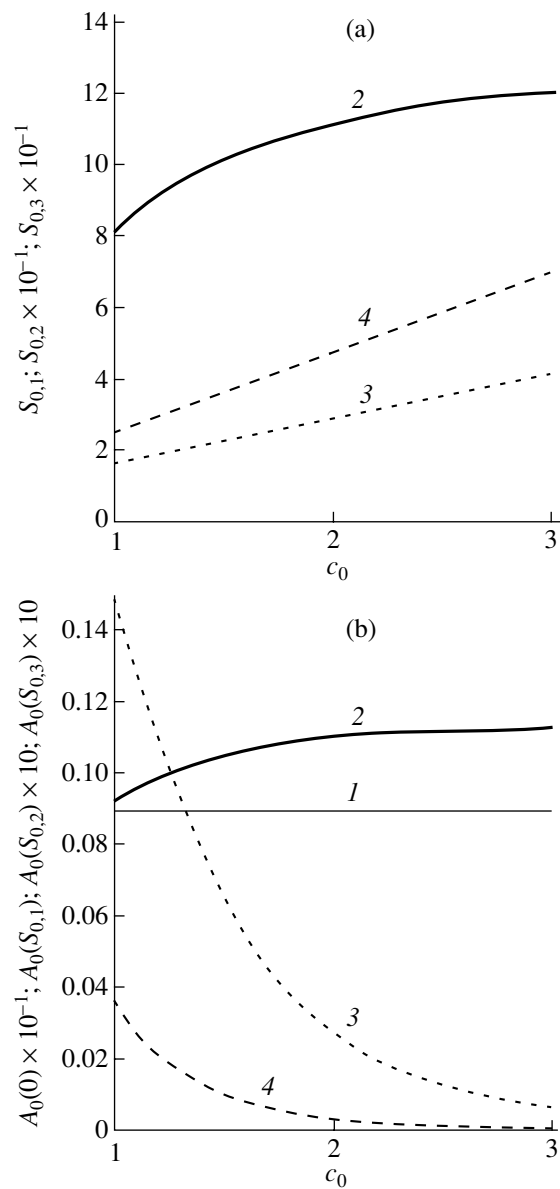

Fig. 3. Plots of (a) the dimensionless frequencies $S_{0,k}$ of radial oscillations of the bubble surface and (b) the amplitude of modes $A_0(S_{0,k})$ vs. dimensionless sound speed c_0 for $P_{0g} = 0.05$, $P_\infty^{(e)} = 0.05$, $W = 0.05$, $\gamma = 4/3$: (*1*) $A_0(0)$; (*2*) $A_0(S_{0,1})$; (*3*) $A_0(S_{0,2})$; and (*4*) $A_0(S_{0,3})$.

$S_{2,k}$ ($k \geq 2$) continue to grow as c_0 increases, while amplitudes $A_2(S_{2,k})$ of deviation of the bubble surface from a spherical shape for these harmonics rapidly tend to zero.

The dependences of the frequencies of the surface oscillations on parameters $P_\infty^{(e)}$, P_{0g}, and W are different for the lowest harmonic ($k = 1$) and harmonics with the numbers $k \geq 2$. For example, the frequencies with numbers $k \geq 2$ (Fig. 4) decrease with increasing P_{0g} and W
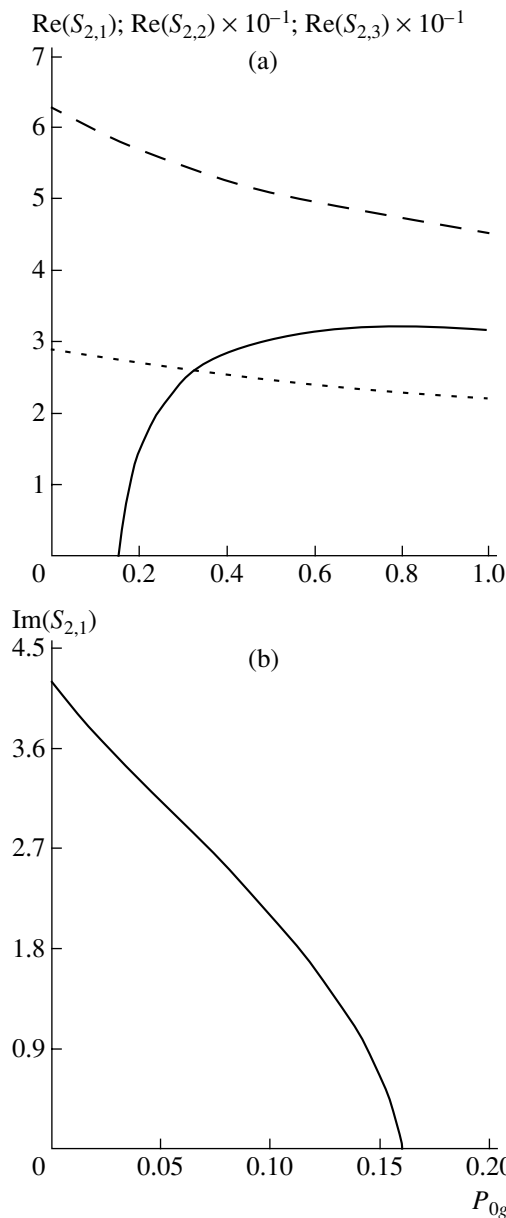
Re($S_{2,1}$); Re($S_{2,2}$) × 10^{-1}; Re($S_{2,3}$) × 10^{-1}

Im($S_{2,1}$)

$M_0(t)$

$M_2(t)$

Fig. 4. Plots of the dimensionless (a) real Re($S_{2,k}$) and (b) imaginary Im($S_{2,k}$) components of the solution of equation $D_2(S) = 0$ determining the frequencies and growth rates of the surface oscillations of a bubble vs. dimensionless gas pressure P_{0g} for $W = 1$, $P_\infty^{(e)} = 1$, $c_0 = 5$, $\gamma = 4/3$.

Fig. 5. Variation of the amplitude coefficients of the (a) radial $M_0(t)$ and (b) surface $M_2(t)$ oscillation modes with time t for $P_{0g} = 0.05$, $W = 0$, $P_\infty^{(e)} = 0.05$, $\gamma = 4/3$, and various magnitudes of the dimensionless sound speed: (thick solid line) $c_0 = 1$; (thin solid line) $c_0 = 5$.

and increase with increasing $P_\infty^{(e)}$. The lowest frequency of the surface oscillations may vanish and become purely imaginary, determining the instability growth rate (this may take place when parameters W or $P_\infty^{(e)}$ increase and also with decreasing P_{0g}). In this case, the further increase of W or $P_\infty^{(e)}$, as well as decrease of P_{0g}, leads to an increase in the growth rate of the instability of the bubble surface (Fig. 4).

The bubble becomes unstable with respect to small perturbations of its shape under condition $\omega_n^2 \leq 0$. This condition for the most unstable second ($n = 2$) mode of the surface oscillations indicates that the instability arises when the electric pressure exceeds the Laplace one, and the gas pressure inside the bubble is leveled with the liquid pressure. This conclusion is physically clear: if the gas pressure in the bubble is higher than the liquid pressure, then the bubble surface perturbation is

smoothed out by the high gas pressure. If the liquid pressure exceeds the gas pressure in the bubble, then any perturbation of the bubble shape is enhanced by the higher liquid pressure, which results in the loss of stability of the bubble shape. The criterion of the loss of stability of the bubble shape can be obtained from conditions $\omega_2^2 \leq 0$ and $P(r) = 0$, which give $W \geq 4\chi\sqrt{P_{0g}/P_\infty^{(e)}}$.

In the region of physical parameters where the bubble may have two equilibrium states (i.e., Eq. (13) has two solutions) the frequencies of the radial and surface (lateral) oscillations that are determined from the solutions of equations $D_n(S) = 0$ ($n \in (\Omega \cup \{0\})$) for the equilibrium state corresponding to the least bubble radius behave in the same way as in the case when the bubble has a single equilibrium state. If the bubble is in the equilibrium state corresponding to the largest radius, then equation $D_0(S) = 0$ always has an imaginary solution determining the growth of instability of the bubble radius The magnitude of this growth rate changes insignificantly with varying gas pressure P_{0g}, parameter W, and sound speed c_0, and significantly increases with decreasing liquid pressure $P_\infty^{(e)}$. When this takes place, the amplitude $A_0(S_{0,1})$ of deviation of the bubble radius from its equilibrium value for this growth rate always remains comparable with unity and increases when $P_\infty^{(e)}$ decreases. The real solutions of equation $D_0(S) = 0$ corresponding to the frequencies of the radial oscillations strongly depend on the liquid pressure $P_\infty^{(e)}$ and the sound velocity c_0 and grow with increasing c_0 and decreasing $P_\infty^{(e)}$, while variations of the gas pressure P_{0g} and the parameter W do not lead to appreciable changes of these solutions. The amplitudes of deviation of the bubble radius from equilibrium $A_0(S_{0,k})$ corresponding to harmonics with $k \geq 2$ are less than unity at least by four orders.

Note that, in the given region of the parameters, the equation $D_n(S) = 0$ ($n \geq 2$) always has real solutions corresponding to the frequencies of the bubble surface oscillations. These solutions are practically independent of the gas pressure P_{0g} in the bubble and of the parameter W, and also considerably increase with decreasing liquid pressure $P_\infty^{(e)}$ and increasing sound speed c_0 (except for the basic harmonic, which is independent of the sound speed). At the same time, the amplitude $A_2(S_{2,1})$ of deviation of the bubble surface from spherical shape corresponding to the basic har-

monic always remains comparable with unity, while the amplitudes $A_2(S_{2,k})$ ($k \geq 2$) of higher harmonics are smaller than unity at least by two orders of magnitude.

CONCLUSIONS

Summarizing the above results, we note that making allowance for the finiteness of the rate of leveling of the gas pressure in a bubble occurring in any of the equilibrium states is necessary only in cases when the sound speed in the gas is a small quantity (when the gas density in the bubble is high). It is only in this case that the amplitudes of high harmonics of the radial and surface oscillations become significant and may appreciably change the bubble shape. The maximum values of the amplitudes of the radial and surface oscillations in expression (33) lower with decreasing the sound speed (Fig. 5). This circumstance influences the surface shape of an oscillating bubble. In the case when the sound speed in a bubble is high (at a low gas density), making allowance for the influence of the higher harmonics of both radial and surface oscillations is unnecessary and expression (38), adequately describing the bubble oscillation at the lowest oscillation frequency, can be used. From the above considerations, it is evident that expression (33) should be employed for the analysis of stability of a collapsing bubble in which the gas pressure at the final stage is very high. Expression (38) can be used for the analysis of motion of a bubble executing small oscillations in the vicinity of one of the equilibrium states.

REFERENCES

1. A. N. Zharov and S. O. Shiryaeva, Elektrokhim. Obrab. Met., No. 6, 9 (1999).

2. F. MacIntyre, J. Geophys. Res. **77**, 5211 (1972).

3. M. Khaleeg-ur-Raman and C. P. R. Saunders, Atmos. Res. **26**, 329 (1991).

4. F. Aitken, F. M. J. McCluskey, and A. Denat, J. Fluid Mech. **327**, 373 (1996).

5. M. E. Glinsky, D. S. Baily, A. R. London, *et al.*, Phys. Fluids **13**, 20 (2001).

6. I. I. Ievlev and A. B. Isers, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 6, 101 (1982).

7. S. T. Zavtrak and E. V. Korobko, Zh. Tekh. Fiz. **61** (3), 177 (1991) [Sov. Phys. Tech. Phys. **36**, 359 (1991)].

8. I. V. Pylaeva, O. A. Sinkevich, and P. V. Smirnov, Teplofiz. Vys. Temp. **30**, 367 (1992).

9. A. P. Vasil'ev, Zh. Tekh. Fiz. **73** (1), 35 (2003) [Tech. Phys. **48**, 31 (2003)].

10. A. I. Grigor'ev, A. N. Zharov, V. A. Koromyslov, and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 205 (1998).

11. A. I. Grigor'ev and A. N. Zharov, Zh. Tekh. Fiz. **70** (4), 8 (2000) [Tech. Phys. **45**, 389 (2000)].

12. A. N. Zharov, O. S. Shiryaeva, and A. I. Grigor'ev, Zh. Tekh. Fiz. **70** (6), 37 (2000) [Tech. Phys. **45**, 704 (2000)].

13. I. N. Didenkulov, D. A. Selivanovskiĭ, V. E. Semenov, and I. V. Sokolov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **42**, 183 (1999).

14. Z. C. Feng and L. G. Leal, J. Fluid Mech. **266**, 209 (1994).

15. S. O. Shiryaeva, A. I. Grigor'ev, V. A. Koromyslov, and A. N. Zharov, Zh. Tekh. Fiz. **73** (9), 60 (2003) [Tech. Phys. **48**, 1141 (2003)].

16. A. N. Zharov and A. I. Grigor'ev, Zh. Tekh. Fiz. **71** (11), 12 (2000) [Tech. Phys. **46**, 1358 (2000)].

17. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953; Inostrannaya Literatura, Moscow, 1958), Vol. 1.

18. A. I. Grigor'ev, S. O. Shiryaeva, and V. A. Koromyslov, Zh. Tekh. Fiz. **68** (9), 1 (1998) [Tech. Phys. **43**, 1011 (1998)].

Translated by N. Mende

GASES
AND LIQUIDS

On the Correct Writing of the Law of Conservation of Amount of Substance at the Moving Fluid–Fluid Interface

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

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

e-mail: grig@uniyar.ac.ru

Received April 6, 2004

Abstract—The reasons for the erroneous writing of the substance conservation law at the moving fluid–fluid interface, which is commonly encountered in related publications, are analyzed. A mathematical statement of this law that is valid for any curved surface that has a nonzero curvature in its equilibrium state is derived in terms of vector analysis. The new writing is independent of the coordinates and can be used for analysis of relaxation phenomena associated with nonlinear wave motions. © *2004 MAIK "Nauka/Interperiodica"*.

(1) The effect of surfactants and charge relaxation phenomena on wave motions at the fluid–fluid interface and on the stability of this interface is the subject of much investigation in modern electrohydrodynamics and physical chemistry. The problem of interface (or free surface) motion with allowance for relaxation can be rigorously stated only for a viscous fluid. The fact is that, in this case alone, the contribution of the relaxation motion of the fluid to the boundary condition for the stress tensor tangential component can be compensate for by viscous stresses. At the interface between (or on the free surface of) perfect fluids, the tangential stresses remain uncompensated [1–3]; hence, the problem is impossible to state correctly.

Since the relaxation effects can be taken into account only for viscous fluids, the scope of problems concerning the wave motion of a fluid where the relaxation effect is considered in the linear (in wave amplitude) approximation is limited. This is because, until recently, nonlinear periodic waves on the fluid surface have been considered within the framework of the perfect fluid model. The methods of solving such problems have been developed only in the last few years [4–6]. The problems of nonlinear oscillation of droplets and jets of a viscous fluid, which are intimately related to the charge relaxation effect and the effect of surfactants, have not yet rigorously resolved. As a result, the problems allowing for both effects have been solved to date only in the linear (in wave or oscillation amplitude) approximation.

The rigorous statement of the substance (charge or surfactant) conservation law at the fluid–fluid interface was given in [1–3], where an approximation linear in deviation of the virtual interface from the equilibrium planar surface was used. Unfortunately, this statement was incorrectly generalized for curvilinear surfaces, with the result that the term proportional to the mean curvature of the undisturbed interface (which equals zero for a planar surface) was lacking. Therefore, a large body of publications concerning relaxation problems on curvilinear surfaces with a nonzero curvature of the undisturbed (equilibrium) surface and using the incorrect statement of this law cited erroneous results (see, e.g., [7–17]). In only a few of them, this term was included in the linear-in amplitude approximation, although no comments on its origin were made [18–20]. Since this error is often encountered in the literature, the rigorous derivation of the substance conservation law in the form adequate to the solution of nonlinear problems, such as interaction between relaxation and capillary waves (shape oscillation), seems to be topical. This is the aim of the present work.

(2) Assume that two viscous incompressible immiscible fluids are separated by interface S. The fluid flow velocity field $\mathbf{W}(\mathbf{r}, t)$, which may be associated, for example, with capillary wave motion at interface S, varies in a continuous manner, so that at any time and at any point on interface S it can be represented in the form

$$\mathbf{W}(\mathbf{r}, t) = \mathbf{u}(\mathbf{r}, t)\mathbf{n} + \mathbf{U}(\mathbf{r}, t)\boldsymbol{\tau},$$

where $\mathbf{n}$ and $\boldsymbol{\tau}$ are the unit vectors of the normal and tangent to interface S at a given point.

We also assume that the fluids have finite thermal conductivities and that an electric charge and surfactants are distributed over interface S in such a way that the capillary wave motion will cause the charge and surfactants to relax. To state the problem of capillary wave motion in a mathematical form, it is necessary to write the laws of charge and surfactant conservation at the interface. In other words, it is necessary to write these conservation laws in differential form for any point on interface S.

To simplify the considerations that follow, we note that the error that is involved in the writing of the substance conservation law [7–17] and is to be eliminated

in this work is by no means related to the presence or absence of diffusion terms. Therefore, we will discuss diffusionless writing of the law to simplify the mathematics. The diffusion components appearing in the complete law of conservation are well known [1, 3] and can be added at the final stage of the derivation.

The primary goal of this work is to show that the diffusionless form of the substance conservation law as applied to moving curvilinear fluid–fluid interface S is given by

$$\frac{\partial \Gamma}{\partial t} + \Gamma u \operatorname{div}(\mathbf{n}) + \operatorname{div}_S(\Gamma U \boldsymbol{\tau}) = 0. \tag{1}$$

Here, Γ is the surface concentration of the substance (charge or surfactants) and div_S is the surface divergence operator.

In [1–3], the conservation law is written in coordinate form and an approximation linear in the amplitude of the velocity field is used. This field is assumed to be associated with capillary wave motion of infinitely small amplitude (hence, the field is also assumed to be small). Then, in terms of vector analysis, Eq. (1) is recast as

$$\frac{\partial \Gamma}{\partial t} + \operatorname{div}_S(\Gamma U \boldsymbol{\tau}) = 0. \tag{2}$$

It was noted [2] that form (2) of the substance conservation law applies only to the planar surface of a fluid. In which form the law should be written when the surface is curvilinear remains unclear.

The model problem of determining the rate of change of the substance concentration on the surface of an expanding bubble is a simple and vivid example of why form (2) of the law is inapplicable to a curvilinear (disturbed) surface having a nonzero mean curvature in its equilibrium (undisturbed) state. Let, at initial time instant t, the charge density or surfactant concentration on the surface of a bubble of radius R be Γ. The bubble expands with constant radial velocity $u \equiv dR/dt$. Then, the charge density (surfactant concentration), which is defined as the amount of charge (surfactant weight) divided by the surface area of the bubble (i.e., $\Gamma = M/(4\pi R^2)$), varies with rate $d\Gamma/dt = -2(dR/dt)M/(4\pi R^3) = -2u\Gamma/R$. It is easy to check that formula (2) gives the absurd result: $d\Gamma/dt = 0$, while the result of formula (1) is correct: $d\Gamma/dt = -2u\Gamma/R$.

This example illustrates that formula (2) is invalid if the problem involves the mechanism of changing the surface geometry (some of its parts are extended or contracted). Term $\Gamma u \operatorname{div}(\mathbf{n})$, by which formula (1) differs from formula (2), is responsible for the variation of the surface concentration due to local extensions or contractions. These distortions take place at the interface when it moves or deforms.

In the above example, the spherical interface moves owing to a change in the volume of the bubble. A wrong impression may arise that, when studying small oscilla-

tions of incompressible droplets or jets, one may get by with the conservation law in form (2) (as in the case of small-amplitude waves on the planar surface of a fluid [1–3]), since the volume bounded by the surface does not change. It is this error that was made in [7–17]. Actually, the curvature of the surface in its equilibrium state, rather than the constancy of the volume, is the crucial factor. Below, we will show that the conservation law should be written in form (1) even if the oscillation amplitude of the droplets and jets is small. Our consideration will follow the scheme with which the formula for the surface tension force pressure under the fluid free surface distorted by virtual deformation was derived [2].

(3) We will start from two rules of vector analysis and differential geometry, which will be used further in derivation of the substance conservation law in form (1).

The first one is the well-known formula for calculating the total time derivative of an integral taken over finite deformable volume $V(t)$ that moves together with the fluid and is bounded by surface ω [21, §120; 22, Chap. 2, Sect. 4]:

$$\frac{d}{dt}\iiint\limits_{V(t)} C(\mathbf{r}, t)dV = \iiint\limits_{V(t)} \frac{\partial C(\mathbf{r}, t)}{\partial t}dV$$

$$+ \iint\limits_{\omega(t)} C(\mathbf{r}, t) \cdot \mathrm{W}(\mathbf{r}, t)d\omega \tag{3}$$

$$\equiv \iiint\limits_{V(t)} \frac{\partial C(\mathbf{r}, t)}{dt}dV + \iint\limits_{\omega(t)} C(\mathbf{r}, t)(\mathbf{W} \cdot \mathbf{N})d\omega.$$

Here, $\mathbf{r}$ is the radius vector of a point in space in the inertial reference system, $\mathrm{W}(\mathbf{r}, t)$ is the velocity field in the fluid, $(\mathbf{W} \cdot \mathbf{N})$ is the algebraic projection of velocity field $\mathbf{W}$ onto normal $\mathbf{N}$ outer to closed surface $\omega(t)$ bounding material volume $V(t)$, and $C(\mathbf{r}, t)$ is the parameter that characterizes some physical property of a small moving fluid particle. Parameter $C(\mathbf{r}, t)$ may be, for example, the local impurity concentration, temperature, energy, etc. The meaning of $V(t)$ on the right and left of formula (3) is different. On the left, $V(t)$ means the material volume moving together with the fluid. The flux of field $\mathbf{W}(\mathbf{r}, t)$ through its boundary, which is constituted by fluid particles, is absent, since the boundary moves together with the particles. On the right, $V(t)$ is viewed as a stationary volume through which the fluid flows. At time instant t under consideration, the geometric boundaries of both volumes coincide, forming surface $\omega(t)$. Designation $\omega(t)$ corresponds to the stationary surface that coincides (at time instant t) with the material surface, which moves together with fluid particles located on the surface bounding material volume $V(t)$.

The second rule is the rule of transformation of slightly deformed surface S into new (virtual) nearby surface S_1. This is done by displacing each point $A \in S$

along the normal issuing out of this point (Fig. 1) by distance $h = h(A)$ (that is, h depends on the position of point A). In our case, S is a separated (finite) part of the interface.

Let the radius vector of point A on surface S be expressed through parameters α and β, which are the curvilinear coordinates of the point, as $\mathbf{r} = \mathbf{r}(\alpha, \beta)$. Then, the normal to surface S at point A can be expressed through the same parameters: $\mathbf{n} = \mathbf{n}(\alpha, \beta)$. In this case, the radius vector $\mathbf{r}_1$ of a point on surface S_1, which is constructed by displacing each point on surface S by distance h along the normal, may be considered dependent on the same parameters α and β: $\mathbf{r}_1 = \mathbf{r} + h \cdot \mathbf{n}$, where $h = h(\alpha, \beta)$ is the function of the curvilinear coordinates of point $A \in S$.

The difference δS between surface S and S_1 is given by the well-known relationship (see, e.g., 21, §139)

$$\delta S = S_1 - S = \iint\limits_{S} n \operatorname{div}(\mathbf{n}) \cdot dS + o(h). \qquad (4)$$

Here, the normal to surface S is designated by $\mathbf{n}$ instead of $\mathbf{N}$, as in the consideration of the former rule, to make a distinction between normal $\mathbf{N}$ to surface ω and normal $\mathbf{n}$ to surface S.

In [21], rule (4) was written in terms of the mean curvature of surface H, which was calculated at point $A \in S$ (Fig. 1). In our case, (4) takes into account the relationship $2H = -\operatorname{div}(\mathbf{n})$, which obviously follows from the well-known Rodrigo theorem of partial derivatives of a normal unit vector along principle curvatures of a surface [23].

From (4), it follows that the smaller h, the more accurate the approximate relationship between elementary surface areas dS and dS_1:

$$dS_1 \approx (1 + h \cdot \operatorname{div}(\mathbf{n}))dS. \qquad (5)$$

(4) Let us derive the substance conservation law on moving curved surface S that is a part of the interface between two media when surface diffusion is absent. For simplicity, we will speak of a surfactant that is distributed over fluid–fluid interface S with surface concentration $\Gamma = \Gamma(A, t)$, where $A \in S$ (all the considerations that follow are also valid if Γ is the surface charge density).

Assume that function $\Gamma = \Gamma(A, t)$ has all partial derivatives with respect to all the arguments and that interface S is simple closed curve L (Fig. 2) that is sufficiently smooth for the formulas of vector analysis to be applied. We also assume that the continuity condition for hydrodynamic velocity field $\mathbf{W}(\mathbf{r}, t)$ is met at the interface.

(i) Surfactant film as the ultimate state of a surfactant layer whose thickness tends to zero. Surface S with the distributed surfactant will be considered as the ultimate state of a surfactant layer with finite thickness h. This layer is obtained as follows. Surface S_1 is constructed by displacing all points of initial surface S
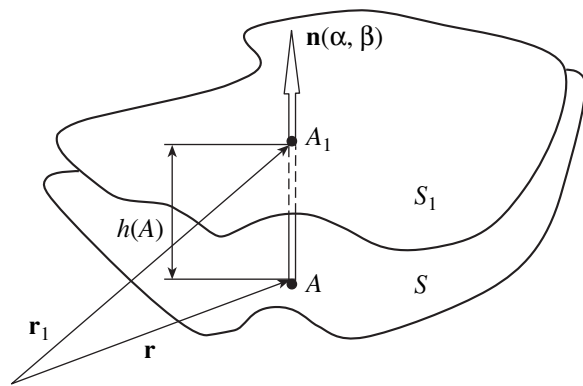

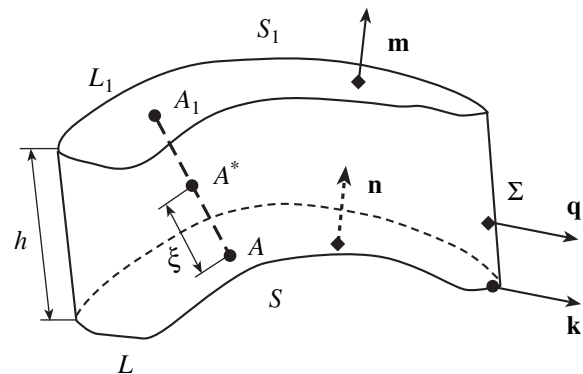**Fig. 1.** Part of interface S and virtual surface S_1.

Fig. 2. Schematic of the h layer.

(Fig. 2) by distance h along the normal (for simplicity, distance h is assumed to be constant, i.e., independent of the curvilinear coordinates of point $A \in S$). The layer bounded by S and S_1 will be called the h layer and designated as h. The value of h is taken to be sufficiently small, so that formulas (4) and (5) can be applied. Contour L_1, which bounds surface S_1, is formed by displacing the points of contour L by distance h along the normal (Fig. 2). The lateral area of the h layer will be designated as Σ.

Generally, the volume concentration C_h in the h layer depends on the position of point $A^* \in \Pi_h$ inside the h layer. This position, in turn, is totally specified by the position of point $A \in S$ and length ξ of segment AA^*, which is normal to S (Fig. 2). Also, let C_h be a function of time t. The model of filling the h layer with a substance of volume concentration $C_h = C_h(A, \xi, t)$ ($A \in S$; $0 \le \xi \le h$) is as follows: it is required that, at $h \longrightarrow 0$, a family of functions $C_h = C_h(A, \xi, t)$ converge uniformly with respect to the position of point A and times $t > 0$:

$$hC_h(A, \xi, t) \Longrightarrow \Gamma(A, t); \quad (h \longrightarrow 0, A \in S, t > 0). \quad (6)$$

Condition (6) imparts a property to the h layer that is crucial for further discussion: the surfactant-filled h layer shrinks, tending to surface S, at $h \longrightarrow 0$, and the
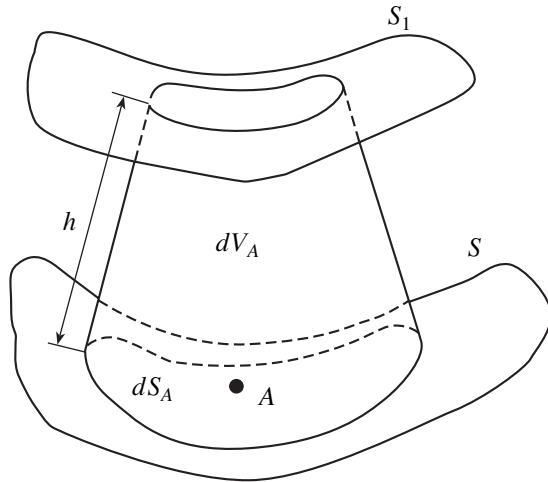
Fig. 3. Elementary volume inside the h layer.

surfactant concentration on this surface will be described by the surface function $\Gamma = \Gamma(A, t)$, where $A \in S$.

Indeed, the thinner the h layer, the more accurately the expression $C_h(A)dV_A = hC_h(A)dS_A$ describes the amount of the substance in volume $dV_A = dS_A h$, which rests on elementary area $dS_A \in S$ containing point A (Fig. 3). However, according to (6), the amount of the substance tends to $\Gamma(A)dS_A$ at $h \longrightarrow 0$, i.e., to the amount of the surfactant that is uniformly distributed over elementary area dS_A.

Uniform convergence in (6) provides the correctness of the passages to the limit that are used below.

(ii) Law of conservation of amount of surfactant on surface S in integral form. Let the h layer move together with the fluid. Then, the fluid flow and, accordingly, the surfactant transfer across the boundary of this layer, $\omega \equiv S \cup \Sigma \cup S_1$, are absent. If the amount of the surfactant in the h layer is M, the surfactant conservation condition in it can be written, in view of (3), in the form

$$\frac{dM}{dt} = \frac{d}{dt}\left(\int\int_{V(t)}\int C_h dV\right) \equiv \int\int_{V(t)}\int \frac{\partial C_h}{\partial t}dV$$

$$+ \int_{S \cup \Sigma \cup S_1}\int C_h(\mathbf{W} \cdot \mathbf{N})d\omega = 0,$$

where

$$\mathbf{N} = \begin{cases} -\mathbf{n} & \text{on} \quad S \\ \mathbf{m} & \text{on} \quad S_1 \\ \mathbf{q} & \text{on} \quad \Sigma. \end{cases}$$

The above definition of vector $\mathbf{N}$ means that it is the outer normal to the h layer volume, which is bounded

by surfaces S, S_1, and Σ. Let the unit vectors of normals $\mathbf{n}$, $\mathbf{m}$, and $\mathbf{q}$ to surfaces $S = S(t)$, $S_1 = S_1(t)$, and $\Sigma = \Sigma(t)$ be directed as shown in Fig. 2. Then, the substance conservation condition in the h layer may be recast as

$$\int\int_{V(t)}\int \frac{\partial C_h}{\partial t}dV + \int_{\Sigma(t)}\int C_h(\mathbf{W} \cdot \mathbf{q})d\Sigma$$

$$+ \left(\int_{S_1(t)}\int C_h(\mathbf{W} \cdot \mathbf{m})dS - \int_{S(t)}\int C_h(\mathbf{W} \cdot \mathbf{n})dS\right) = 0. \tag{7}$$

It is easy to find the limiting form of expression (7) at $h \longrightarrow 0$. To this end, we make use of the integral theorem of mean, relationship (6), and expression (5) for the parenthesized terms in (7). The passage to the limit for the terms appearing in (7) is described in detail in the appendix. Eventually, (7) is transformed into the integral form of the substance conservation law on surface S:

$$\int_{S(t)}\int \frac{\partial \Gamma}{\partial t}dS + \oint_{L(t)} \Gamma(\mathbf{W} \cdot \mathbf{k})dL + \int_{S(t)}\int \Gamma u \,\text{div}(\mathbf{n})dS = 0. \tag{8}$$

This law for the surfactant is fulfilled on any finite surface S bounded by contour L, provided that this surface is a part of the moving fluid–fluid interface at a time t. In (8), $\mathbf{k}$ is the unit vector that is tangent to surface S at point $\Lambda \in L$. This unit vector is perpendicular to contour L at point Λ, which is coincident with the extremity of unit vector $\mathbf{q}$. In turn, $\mathbf{q}$ is the unit vector of the outer normal to the lateral surface of the h layer at $h \longrightarrow 0$ (Fig. 4).

Relationship (8), which is the integral form of the law of conservation of amount of substance on the moving fluid–fluid interface, is of independent interest for physicochemical hydrodynamics.

(iii) Law of conservation of amount of surfactant on surface S in differential form. Dividing both sides of (8) by S yields an expression that is valid for any part of the moving fluid–fluid interface. Therefore, it will remain valid if surface S is shrunk to point $A \in S$. In this case, using the integral theorem of mean, we can recast the first and last terms on the left of (8) as

$$\frac{1}{S}\int_{S(t)}\int \frac{\partial \Gamma}{\partial t}dS \longrightarrow \left(\frac{\partial \Gamma}{\partial t}\Big|_{A_0 \in S}\right)\frac{1}{S}\int_{S(t)}\int dS \longrightarrow \frac{\partial \Gamma}{\partial t}\Big|_A; \tag{9}$$

$$\frac{1}{S}\int_{S(t)}\int \Gamma u \,\text{div}(\mathbf{n})dS \longrightarrow ((\Gamma u \,\text{div}(\mathbf{n}))|_{A_0 \in S})$$

$$\times \frac{1}{S}\int_{S(t)}\int dS \longrightarrow (\Gamma u \,\text{div}(\mathbf{n}))|_A. \tag{10}$$

Here, point A_0 is a point on surface S ($A_0 \in S$) that is specified for each of the integrals based on the theorem of mean. The position of point A_0 on surface S is of no concern, since $A_0 \longrightarrow A$ when contour L shrinks to

point A. Dividing the medium term on the left of (8) by S yields (in the limit when S shrinks to point A) an expression for surface divergence at point A:

$$\frac{1}{S}\oint_{L(t)}\Gamma(\mathbf{W}\cdot\mathbf{k})dL$$

$$= \frac{1}{S}\oint_{L(t)}\Gamma U(\boldsymbol{\tau}\cdot\mathbf{k})dL \longrightarrow \mathrm{div}_S((\Gamma u)\boldsymbol{\tau})|_A. \tag{11}$$

Recall that $\mathbf{k}$ is the normal unit vector to contour $L(t)$ and the integral on the left of (11) is the net flux of the vector field $(\Gamma U)\boldsymbol{\tau}$ through boundary $L(t)$. Note also that relationship (11) takes into account that $\mathbf{n}\cdot\mathbf{k}\equiv 0$.

Let us replace the terms on the left of (8) by their limiting formulas (9)–(11) and, taking advantage of the fact that point $A \in S$ was chosen arbitrarily, omit the indication that the calculations are performed at point A. Eventually, we arrive at the substance conservation law at the moving fluid–fluid interface in differential form (1), which was to be proved.

(5) Consider the planar surface of a fluid that occupies (entirely or partially) the half-space $z < 0$ in the coordinate system $0xyz$ with the $0z$ axis directed opposite to the gravitational force. If waves start propagating on this surface in the $0x$ direction, the tangent and normal unit vectors on the disturbed (offset from its equilibrium position) surface $z = \xi(x, t)$ have the form

$$\mathbf{n} = -\frac{\partial_x\xi}{\sqrt{1+(\partial_x\xi)^2}}\mathbf{e}_x + \frac{1}{\sqrt{1+(\partial_x\xi)^2}}\mathbf{e}_z;$$

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

Here, $\mathbf{e}_x$ and $\mathbf{e}_z$ are the unit vectors along the Cartesian axes. If, for simplicity, the motion of the fluid is assumed to be independent of the coordinate y, $\mathbf{U} = w\mathbf{e}_x + v\mathbf{e}_z$ and condition (1) in coordinate form can be written as

$$z = \xi: \frac{\partial\Gamma}{\partial t}$$

$$+ \frac{1}{1+(\partial_x\xi)^2}(\partial_x(\Gamma w) + \partial_x\xi(\Gamma\partial_z w + \partial_x(v\Gamma)) \tag{12}$$

$$+ (\partial_x\xi)^2\Gamma\partial_z v) + D = 0.$$

Here, D stands for diffusion terms previously omitted. This form of the conservation law was used in [19]. In the approximation linear in wave amplitude, we, with regard to (12), come to the expression used in [1, 2]:

$$z = 0: \frac{\partial}{\partial t} + \partial_x(\Gamma w) + D = 0.$$

(6) Let us discuss the error in the writing of the substance conservation law on the surface of droplets and
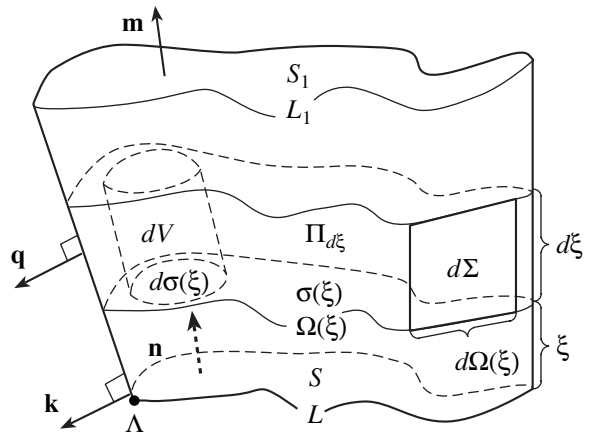
Fig. 4. Partition of the h layer into elementary volumes dV_A and of its lateral surface Σ into elementary areas $d\Sigma$.

jets slightly oscillating about their equilibrium shape (for the droplets, the equilibrium shape is a sphere of radius R; for the jets, a cylindrical surface of radius R). In the approximation of small oscillation amplitude, the medium term in (1) is a quantity of the first order of smallness and is given by

$$\Gamma u\mathrm{div}(\mathbf{n}) \approx \frac{2}{R}\Gamma^{(0)}u^{(1)},$$

for a droplet of radius R and

$$\Gamma u\mathrm{div}(\mathbf{n}) \approx \frac{1}{R}\Gamma^{(0)}u^{(1)}$$

for a jet of radius R.

Here, $\Gamma^{(0)}$ is the equilibrium concentration of the substance (charge or surfactant) on the undisturbed spherical or cylindrical surface and $u^{(1)}$ is the radial velocity of the fluid in the first approximation in oscillation amplitude.

For waves on the planar surface, the medium term in (1) calculated in the linear-in-amplitude approximation vanishes because of the zero curvature of the equilibrium planar surface.

CONCLUSIONS

Our derivation of the substance conservation law at the moving fluid–fluid interface indicates that it is necessary to take into account the term that is proportional to the mean curvature of the interface. This term is often erroneously disregarded when relaxation motions at the curvilinear interface are calculated in an approximation linear in deviation of the surface from its equilibrium shape. In the higher order calculations, this term should be included not only for the curvilinear interface but also for the planar surface.

In this work, the substance conservation law at the moving fluid–fluid interface was used in the coordi-

nateless form (it was derived in terms of vector analysis). This form of writing is appropriate for the statement and solution of the problems that are concerned with the influence of charge or surfactant relaxation on the evolution of the fluid–fluid interface. Moreover, this form does not depend on the equilibrium shape of the surface in external force fields and on the degree of its deformation during motion. This circumstance is of special importance for researchers engaged in the construction of nonlinear models of periodic wave motion on the charged surfactant-covered interface between two viscous fluids of finite electrical conductivity.

APPENDIX

Passage to the Limit of an Infinitely Thin h Layer in Relationship (7) for the Substance Conservation Law

(1) The first term on the left of (7). Let the h layer be cut by surface $\sigma = \sigma(\xi)$ that is constructed by displacing all points of surface S by a distance $0 \le \xi \le h$ along the normal to S. For $\xi = 0$, we have $\sigma(0) \equiv S \equiv S(t)$; for $\xi = h$, $\sigma(h) \equiv S_1(t)$. For the layer $\Pi_{d\xi}$, which has thickness $d\xi$ much smaller than h and rests on $\sigma(\xi)$, an elementary volume has the form $dV = d\xi d\sigma(\xi)$ (Fig. 4). Such elementary volumes occupy volume V layerwise, and the integral sum constructed for the first integral on the left of (7) turns into a repeated integral in the limit $dV \longrightarrow 0$:

$$\iiint_{V(t)} \frac{\partial C_h}{\partial t} dV = \int_0^h d\xi \iint_{\sigma(\xi)} \frac{\partial C_h}{\partial t} d\sigma(\xi).$$

According to the integral theorem of mean, there exists $\xi = \xi^* \in [0, h]$ such that

$$\iiint_{V(t)} \frac{\partial C_h}{\partial t} dV = h \iint_{\sigma(\xi^*)} \frac{\partial C_h}{\partial t} d\sigma(\xi^*)$$

$$= \iint_{\sigma(\xi^*)} \frac{\partial (h C_h)}{\partial t} d\sigma(\xi^*).$$

Here, we used the fact that h is assumed to be constant. In the limit $h \longrightarrow 0$, $\xi^* \longrightarrow 0$ and $\sigma(\xi^*) \longrightarrow S(t)$; also, relationship (6) is fulfilled. Therefore, this limit is easy to find:

$$h \longrightarrow 0: \iiint_{V(t)} \frac{\partial C_h}{\partial t} dV \longrightarrow \iint_{S(t)} \frac{\partial \Gamma}{\partial t} dS.$$

(2) The second term on the left of (7). Let the h layer be cut by surfaces $\sigma(\xi)$ introduced previously (Fig. 4) and $\Omega = \Omega(\xi)$ be the contour that is obtained when the lateral surface Σ of the h layer meets surface $\sigma(\xi)$. It is obvious that $\Omega(0) = L$ and $\Omega(h) = L_1$ (Fig. 4). On the lateral surface of the $\Pi_{d\xi}$ layer (see above), we separate an elementary area $d\Sigma = d\xi d\Omega(\xi)$ (Fig. 4). The lateral surfaces of the $\Pi_{d\xi}$ layer, which consist of such elementary

areas, produce bands occupying the entire lateral surface Σ of volume V. Eventually, the lateral surface Σ of the h layer turns out to be partitioned into elementary areas. With this partition, the integral sum for the second integral on the left of (7) turns to a repeated integral in the limit $d\Sigma \longrightarrow 0$:

$$\iint_{S(t)} C_h(\mathbf{W} \cdot \mathbf{q}) d\Sigma = \int_0^h \left(d\xi \oint_{\Omega(\xi)} C_h(\mathbf{W} \cdot \mathbf{q}) \right) d\Omega(\xi).$$

According to the integral theorem of mean, there exists $\xi = \xi^* \in [0, h]$ such that

$$\iint_{S(t)} C_h(\mathbf{W} \cdot \mathbf{q}) d\Sigma = \int_0^h d\xi \oint_{\Omega(\xi)} C_h(\mathbf{W} \cdot \mathbf{q}) d\Omega(\xi)$$

$$= h \oint_{\Omega(\xi^*)} C_h(\mathbf{W} \cdot \mathbf{q}) d\Omega(\xi) = \oint_{\Omega(\xi^*)} h C_h(\mathbf{W} \cdot \mathbf{q}) d\Omega(\xi).$$

In the limit $h \longrightarrow 0$, $\xi^* \longrightarrow 0$ and $\Omega(\xi^*) \longrightarrow L(t)$; also, relationship (6) is fulfilled and we may pass to the limit $\mathbf{q} \longrightarrow \mathbf{k}$, where $\mathbf{k}$ is the unit vector that is tangent to surface S and perpendicular to at point $\Lambda \in L$ considered (Fig. 4). Therefore, this limit is easy to find:

$$h \longrightarrow 0: \iint_{S(t)} C_h(\mathbf{W} \cdot \mathbf{q}) d\Sigma \longrightarrow \oint_{L(t)} \Gamma(\mathbf{W} \cdot \mathbf{k}) dL.$$

(3) In view of (5), the parenthesized terms on the left of (7) can be recast as

$$\lim_{h \to 0} \left(\iint_{S_1(t)} C_h(\mathbf{W} \cdot \mathbf{m}) dS_1 - \iint_{S(t)} C_h(\mathbf{W} \cdot \mathbf{n}) dS \right)$$

$$= \lim_{h \to 0} \left(\iint_{S(t)} C_h^+(\mathbf{W} \cdot \mathbf{m}) dS \right.$$

$$+ \iint_{S(t)} h C_h^+ \mathbf{W} \cdot \mathbf{m} \, \mathrm{div}(\mathbf{n}) dS - \iint_{S(t)} C_h^0(\mathbf{W} \cdot \mathbf{n}) dS \right).$$

Surfactant concentration C_h at inner point A^* of the h layer (Fig. 2) depends on the position of point $A \in S$ and length ξ of segment AA^*, which is normal to S: $C_h = C_h(A, \xi, t)$. In the above expression, $C_h^0 \equiv C_h(A, 0, t)$ and $C_h^+ \equiv C_h(A, h, t)$.

On the right of this expression, the first and last integrals tend to infinity at $h \longrightarrow 0$. To estimate the rate with which these integrals tend to infinity, let us consider the asymptotic behavior of volume surfactant concentration C_h at $h \longrightarrow 0$ (see (6)). For condition (6) to be met, it is necessary that

$$C_h(A, \xi, t) \sim \frac{\Gamma(A, t)}{h} \longrightarrow \infty \quad (h \longrightarrow 0).$$

On the other hand, since ξ in this asymptotic expression is taken arbitrarily, we get the following asymptotic formulas:

$$h \longrightarrow 0: \left.\begin{array}{l} C_h^+ = C_h(A, h, t) \sim \dfrac{\Gamma(A, t)}{h} \\[2mm] C_h^0 = C_h(A, 0, t) \sim \dfrac{\Gamma(A, t)}{h} \end{array}\right\} \Rightarrow C_h^0 \longrightarrow C_h^+.$$

In addition, the continuity of field velocity $\mathbf{U}$ implies that $(\mathbf{W} \cdot \mathbf{m}) \longrightarrow (\mathbf{W} \cdot \mathbf{n})$ at $h \longrightarrow 0$. As a result, the first and last integrals cancel each other at $h \longrightarrow 0$, so that we may write the limiting relationship in the form

$$\lim_{h \to 0} \left(\int\!\!\!\int_{S_1(t)} C_h (\mathbf{W} \cdot \mathbf{m}) dS_1 - \int\!\!\!\int_{S(t)} C_h (\mathbf{W} \cdot \mathbf{n}) dS \right)$$

$$= \lim_{h \to 0} \left(\int\!\!\!\int_{S(t)} h C_h^0 (\mathbf{W} \cdot \mathbf{m}) \mathrm{div}(\mathbf{n}) dS \right).$$

The limit on the left can be found from condition (6), which is met, specifically, for $C_h^0 \equiv C_h(A, 0, t)$. Also, one should take into account that $(\mathbf{W} \cdot \mathbf{m}) \longrightarrow (\mathbf{W} \cdot \mathbf{n}) = U$ at $h \longrightarrow 0$. Eventually, we get

$$\lim_{h \to 0} \left(\int\!\!\!\int_{S_1(t)} C_h (\mathbf{W} \cdot \mathbf{m}) dS_1 - \int\!\!\!\int_{S(t)} C_h (\mathbf{W} \cdot \mathbf{n}) dS \right)$$

$$= \int\!\!\!\int_{S(t)} \Gamma U \mathrm{div}(\mathbf{n}) dS.$$

ACKNOWLEDGMENTS

REFERENCES

1. V. G. Levich, *Physicochemical Hydrodynamics* (Fizmatgiz, Moscow, 1959) [in Russian].

2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).

3. J. R. Melcher, *Field-Coupled Surface Waves* (MIT, Cambridge, 1963).

4. D. F. Belonozhko and A. I. Grigor'ev, Zh. Tekh. Fiz. **73** (4), 28 (2003) [Tech. Phys. **48**, 404 (2003)].

5. D. F. Belonozhko and A. I. Grigor'ev, Zh. Tekh. Fiz. **73** (11), 37 (2003) [Tech. Phys. **48**, 1396 (2003)].

6. D. F. Belonozhko and A. I. Grigor'ev, Zh. Tekh. Fiz. **74** (3), 5 (2004) [Tech. Phys. **49**, 287 (2004)].

7. D. Saville, J. Fluid Mech. **48**, 815 (1971).

8. A. F. Ginevskiĭ and A. I. Motin, Inzh.-Fiz. Zh. **60**, 576 (1991).

9. A. J. Mestel, J. Fluid Mech. **274**, 93 (1994).

10. A. Gonzalez, A. Castellanos, and M. G. Velarde, in *Series on Nonlinear Sciences B*, Vol. 5: *Fluid Physics*, Ed. by M. G. Velarde and C. I. Christov (World Sci., London, 1994), pp. 442–459.

11. A. J. Mestel, J. Fluid Mech. **312**, 311 (1996).

12. V. Ya. Shkadov and A. A. Shutov, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 2, 29 (1998).

13. A. A. Shutov and V. Ya. Shkadov, in *Proceedings of the Regional Contest of Projects in Natural Sciences* (Éĭdos, Kaluga, 2000), Vol. 1, pp. 67–88.

14. Jean-Marc Chomaz, J. Fluid Mech. **442**, 387 (2001).

15. V. Ya. Shkadov and A. A. Shutov, Fluid Dyn. Res. **28**, 23 (2001).

16. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 74 (2002).

17. A. I. Grigor'ev, S. O. Shiryaeva, and V. A. Koromyslov, Zh. Tekh. Fiz. **72** (6), 19 (2002) [Tech. Phys. **47**, 673 (2002)].

18. Hui-Lan Lu and R. E. Apfel, J. Fluid Mech. **222**, 351 (1991).

19. Wu-Ting Tsai and K. P. Dick, J. Fluid Mech. **289**, 315 (1995).

20. Yu. K. Bratukhin, S. R. Kosvintsev, and S. O. Makarov, Kolloidn. Zh. **63**, 259 (2001).

21. V. I. Smirnov, *A Course of Higher Mathematics* (Nauka, Moscow, 1965; Addison-Wesley, Reading, 1964), Vol. 2.

22. J. Serrin, in *Handbuch der Physik*, Vol. 8: *Mathematical Principles of Classical Fluid Mechanics* (Springer-Verlag, Berlin, 1959; NITs "Regul. Khaot. Dinamika," Izhevsk, 2001), pp. 125–263.

23. *Mathematical Encyclopedia*, Ed. by I. M. Vinogradov (Sov. Éntsiklopediya, Moscow, 1984), Vol. 4 [in Russian].

Translated by V. Isaakyan

GASES
AND LIQUIDS

Nonlinear Periodic Waves on the Charged Surfactant-Covered Surface of a Viscous Fluid

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

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

e-mail: grig@uniyar.ac.ru

Received April 6, 2004

Abstract—The profile of a periodic capillary–gravitational wave propagating over the surfactant-covered surface of a fluid is found in the second-order approximation in initial deformation amplitude. It is shown that the surfactant film appreciably affects the intensity of nonlinear interaction between harmonics constituting the nonlinear wave. © 2004 MAIK "Nauka/Interperiodica".

The authors have recently published a number of articles devoted to the asymptotic behavior of nonlinear capillary–gravitational waves over the surface of a finite-viscosity charged fluid [1–4]. The apparatus developed in those articles makes it possible to rigorously take into account viscosity in the problems of nonlinear wave motion. As a result, there has appeared the possibility of studying relaxation phenomena associated with nonlinear wave motion. In other words, there appeared the possibility of adequately considering the balance of viscous and relaxation stresses on the free surface of the fluid in nonlinear problems. In this work, we study the effect of a surfactant film on the intensity of interaction between harmonics constituting a nonlinear periodic wave propagating over the surfactant-covered charged surface of the fluid.

The effect of surfactants on wave propagation over the fluid surface is the subject matter of classical mechanics of fluid and physical chemistry [5]. This issue is of both scientific and applied interest [6–11]. To date, all the related investigations have been carried out in the approximation of infinitely small wave amplitude and have been devoted primarily to the determination of the surfactant concentration that effectively suppresses capillary–gravitational waves. The existence of such an optimal concentration has been proved experimentally (see, e.g., [6] and Refs. therein) and predicted theoretically in terms of various models [6, 7, 9, 11]. Yet, sound physical explanations of this effect are still lacking, and this issue is addressed now and then [9, 11]. In this work, we break with this tradition with the aim of revealing the surfactant effects that are unrelated to the decay of capillary wave motion.

1. PROBLEM DEFINITION

Let a viscous incompressible perfectly conducting fluid of density ρ and kinematic viscosity ν occupy the half-space $z < 0$ in the coordinate system $Oxyz$ with the Oz axis directed oppositely to the gravitational force. An electric charge with surface density κ_0 and a surfactant with surface concentration Γ_0 are uniformly distributed over the free surface of the fluid. Let D be the free diffusion coefficient of the surfactant. We assume that a periodic wave of wavelength λ starts traveling over the free surface in the positive Ox direction at zero time ($t = 0$). Our goal is to find the profile of the wave at $t > 0$.

The periodic profile of the wave can be uniquely restored from the amplitudes of harmonics appearing in the expansion into the Fourier series over spatial period λ. Let the amplitude η of the fundamental harmonic be known. Below, we will use wavenumber $k = 2\pi/\lambda$ instead of the wavelength. For simplicity, the motion of the fluid is assumed to be independent of coordinate y.

As the wave propagates, the distribution of the surfactant over the free surface changes continuously and its concentration becomes a function of time and horizontal coordinate: $\Gamma = \Gamma(t, x)$. Local changes in the surfactant concentration cause local changes in surface tension coefficient γ. As for the dependence $\gamma = \gamma(\Gamma)$, we assume that the surface phase of the surfactant and the fluid are in local thermodynamic equilibrium. This means that a local change in the surfactant concentration instantly causes a change in the surface tension coefficient according to the isotherm $\gamma = \gamma(\Gamma)$ that is deemed to be known.

Mathematically, the problem of determining the wave profile is stated as follows:

$$\partial_t \mathbf{U} + (\mathbf{U} \cdot \boldsymbol{\nabla})\mathbf{U} = -\frac{1}{\rho}\boldsymbol{\nabla}p + \nu\Delta\mathbf{U} + \mathbf{g};$$

$$\mathbf{U} = u\mathbf{e}_x + v\mathbf{e}_z;$$

$$\boldsymbol{\nabla} \cdot \mathbf{U} = 0; \quad \Delta\Phi = 0;$$

$$z = \xi: \partial_t\xi + u\partial_x\xi = v;$$

$$p - 2\rho\nu(\mathbf{n}(\mathbf{n}\cdot\nabla)\mathbf{U}) + \frac{(\nabla\Phi)^2}{8\pi} = -\frac{\gamma\partial_{xx}\xi}{(1 + (\partial_x\xi)^2)^{3/2}};$$

$$-\rho\nu[(\tau(\mathbf{n}\cdot\nabla)\mathbf{U}) + (\mathbf{n}(\tau\cdot\nabla)\mathbf{U})]$$

$$= \frac{\partial_x\gamma}{\sqrt{1 + (\partial_x\xi)^2}} = 0; \quad \Phi = 0;$$

$$\partial_1\Gamma + \frac{1}{1 + (\partial_x\xi)^2}[\partial_x(\Gamma u) + \partial_x\xi(\Gamma\partial_z u + \partial_x(\Gamma v))$$

$$+ (\partial_x\xi)^2\Gamma\partial_z v] - D\left(\frac{\partial_{xx}\Gamma}{1 + (\partial_x\xi)^2} - \frac{\partial_x\xi\partial_{xx}\xi\partial_x\Gamma}{(1 + (\partial_x\xi)^2)^2}\right) = 0;$$

$$z \longrightarrow -\infty: u \longrightarrow 0; \quad v \longrightarrow 0;$$

$$z \longrightarrow \infty; \quad \nabla\Phi \longrightarrow E_0\mathbf{e}_z; \quad E_0 = 4\pi\kappa_0.$$

Here, $\mathbf{e}_x$ and $\mathbf{e}_z$ are the unit vectors along the axes, $\mathbf{n}$ and τ are the unit vectors of the outer normal and tangent to the free surface $z = \xi \equiv \xi(t, x, z)$ disturbed by the wave motion (analytical expressions for $\mathbf{n}$ and τ are given in the Appendix), and Δ is the Laplacian. The boundary conditions for the problem will be set in the course of solution so as to obtain an expression for the wave profile that is as simple as possible and suitable for qualitative analysis (the routine procedure in solving the problems of nonlinear periodic waves [1–4, 12, 13]).

In the rigorous statement, the input data are η, k, ρ, g, ν, κ_0, Γ_0, D, and the isotherm $\gamma = \gamma(\Gamma)$. The unknown functions are $\xi = \xi(t, x, z)$, the profile of the free surface; $u = u(t, x, z)$ and $v = v(t, x, z)$, the horizontal and vertical components of velocity field $\mathbf{U}$ in the fluid; $p = p(t, x, z)$, the pressure distribution in the liquid; $\Gamma = \Gamma(t, x)$, the surface concentration of the surfactant; and $\Phi = \Phi(t, x, z)$, the electric field potential above the fluid.

2. CONSTRUCTION OF ASYMPTOTIC SOLUTION

According to the technique used earlier to solve the problem of nonlinear waves in viscous fluid in the absence of surfactant [3, 4], a solution is sought in the form

$$\begin{pmatrix}\xi\\u\\v\\p\\\Phi\\\Gamma\end{pmatrix} = \begin{pmatrix}0\\0\\0\\-\rho gz - \frac{E_0^2}{8\pi}\\E_0 z\\\Gamma_0\end{pmatrix} + \begin{pmatrix}\xi_1\\u_1\\v_1\\p_1\\\Phi_1\\\Gamma_1\end{pmatrix} + \begin{pmatrix}\xi_2\\u_2\\v_2\\p_2\\\Phi_2\\\Gamma_2\end{pmatrix} + \begin{pmatrix}O(\xi_1^3)\\O(u_1^3)\\O(v_1^3)\\O(p_1^3)\\O(\Phi_1^3)\\O(\Gamma_1^3)\end{pmatrix};$$

$$\begin{pmatrix}\xi_2\\u_2\\v_2\\p_2\\\Phi_2\\\Gamma_2\end{pmatrix} = \begin{pmatrix}O(\xi_1^2)\\O(u_1^2)\\O(v_1^2)\\O(p_1^2)\\O(\Phi_1^2)\\O(\Gamma_1^2)\end{pmatrix};$$

$$\xi_1 = \eta f(t)\cos(kx - \omega t); \quad f(0) = 1.$$

Here, the quantities with subscript 1 refer to the linear (in amplitude) approximation and those with subscript 2 are second-order corrections.

The complete statement of the problem includes surface tension coefficient $\gamma = \gamma(\Gamma)$ and its partial derivatives with respect to x. From the power series

$$\gamma = \gamma(\Gamma) = \gamma_0 + (\Gamma_1 + \Gamma_2 + \ldots)(\partial_\Gamma\gamma)_0$$

$$+ (\Gamma_1 + \Gamma_2 + \ldots)^2(\partial_{\Gamma\Gamma}\gamma)_0 + \ldots;$$

$$\partial_x\gamma = (\partial_\Gamma\gamma)(\partial_x\Gamma)$$

$$= ((\partial_\Gamma\gamma)_0 + (\Gamma_1 + \Gamma_2 + \ldots)(\partial_{\Gamma\Gamma}\gamma)_0 + \ldots)(\partial_x\Gamma),$$

where subscript 0 implies that the function is calculated on the undisturbed (planar) surface of the fluid, we have, at $\Gamma = \Gamma_0$,

$$\gamma\partial_{xx}\xi \approx \gamma_0\partial_{xx}\xi_1 + \gamma_0\partial_{xx}\xi_2 + \chi\Gamma_1\partial_{xx}\xi_1;$$

$$\partial_x\gamma \approx \chi\partial_x\Gamma_1 + \chi\partial_x\Gamma_2 + \beta\Gamma_1\partial_x\Gamma_1$$

accurate to the second order of smallness, where $\chi \equiv (\partial_\Gamma\gamma)_0$ and $\beta \equiv (\partial_{\gamma\gamma}\gamma)_0$.

Using these approximate relationships, as well as the asymptotic expressions for ξ, u, v, p, Φ, and Γ, one can easily pose the first- and second-order problems by the technique used in [3, 4]. The complete mathematical statement of the first- and second-order problems has the form

$$\partial_1\mathbf{U}_m + \frac{1}{\rho}\nabla p_m - \nu\Delta\mathbf{U}_m = \mathbf{V}_m;$$

$$\nabla\cdot\mathbf{U}_m = 0; \quad \Delta\Phi_m = 0;$$

$$z = \xi: \partial_t\xi_m - v_m = f_{1m};$$

$$p_m - 2\rho\nu\partial_z v_m - \frac{E_0}{4\pi}\partial_z\Phi_m + \gamma_0\partial_{xx}\xi_m = f_{2m};$$

$$-\rho\nu(\partial_z u_m + \partial_x v_m) + \chi\partial_x\Gamma_m = f_{3m}; \quad \Phi_m - E_0\xi_m = f_{4m};$$

$$\partial_1\Gamma_m + \Gamma_0\partial_x u_m - D\partial_{xx}\Gamma_m = f_{5m};$$

$$z \longrightarrow -\infty: u_m \longrightarrow 0; \quad v_m \longrightarrow 0;$$

$$z \longrightarrow \infty: |\nabla\Phi_m| \longrightarrow 0.$$

For $m = 1$ and 2, we are dealing with the first- and second-order problems. For the former, $\mathbf{V}_1 = 0$ and $f_{n1} = 0$ ($n = 1$–5). For the latter, $\mathbf{V}_2$ and $f_{n2} = 0$ ($n = 1$–5) are expressed via solutions to the first-order problem by the formulas given in the appendix.

3. SOLUTION OF THE PROBLEM IN THE SECOND-ORDER APPROXIMATION IN PERIODIC TRAVELING WAVE AMPLITUDE

By solving the first- and second-order problems in succession [3, 4], it is easy to find an expression for the profile of a periodic travelling capillary–gravitational wave in second-order approximation in amplitude η:

$$\xi = \eta\cos\theta\exp(rt)$$
$$+ 2\eta^2[\operatorname{Re}(\zeta)\cos(2\theta) - \operatorname{Im}(\zeta)\sin(2\theta)]\exp(2rt); \quad (1)$$

$$\theta = \omega t - kx; \quad \omega = \operatorname{Im}(S); \quad r = \operatorname{Re}(S); \quad \zeta = \frac{M_1}{M_0}.$$

Here, S is the complex frequency that is calculated in the first-order problem, while parameters M_0 and M_1 are calculated in the course of solution of the second-order problem (see Appendix). Owing to a large array of input data and intermediate variables, it is possible to construct the functional dependence of S, M_0, and M_1 on the input data of the problem. Let us construct numerical parameters

$$\Pi_0 = \gamma(\Gamma_0); \quad \Pi_1 = \Gamma_0\left(\frac{d\gamma}{d\Gamma}\right)_{\Gamma = \Gamma_0};$$
$$\Pi_2 = \Gamma_0^2\left(\frac{d^2\gamma}{d\Gamma^2}\right)_{\Gamma = \Gamma_0}, \quad (2)$$

using function $\gamma = \gamma(\Gamma)$.

Parameter$_0$, having the meaning of the surface tension coefficient on the equilibrium (planar) free surface of the fluid, can be used to convert parameter κ_0 to dimensionless Tonks–Frenkel parameter [8, 14]

$$W = 4\pi\kappa_0^2/\sqrt{\rho g\Pi_0}. \quad (3)$$

Now, we will show that parameters $S = S(\rho, g, \nu, k, W, D, \Pi_0, \Pi_1)$, $M_0 = M_0(\rho, g, \nu, k, W, D, \Pi_0, \Pi_1)$, and $M_1 = M_1(\rho, g, \nu, k, W, D, \Pi_0, \Pi_1, \Pi_2)$ can be directly expressed through parameters ρ, g, ν, k, W, D, Π_0, Π_1, and Π_2. Note that only M_1 depends on Π_2.

The complex frequency is calculated by the formula

$$S = \omega_0\alpha(\rho, g, \nu, k, W, D, \Pi_0, \Pi_1); \quad (4)$$

where

$$\omega_0^2 = kg(1 + (ak)^2 - akW); \quad a = \sqrt{\frac{\Pi_0}{\rho g}}$$

a is the capillary constant, and dimensional complex parameter α is that root of the dimensionless dispersion

relation corresponding to the capillary–gravitational wave (the dispersion relation and the way of choosing the appropriate root are given in the appendix). Parameter ω_0, which is intermediate in the calculations, represents the frequency of gravitational–capillary waves that have an infinitely small amplitude and wavenumber k on the surface of a perfect fluid with constant surface tension coefficient Π_0.

The set of parameters μ, ρ, g, ν, k, W, D, Π_0, Π_1, and Π_2, which specify the profile of the nonlinear wave by formula (1), has two important properties. First, it lacks the isotherm. Instead of function $\gamma = \gamma(\Gamma)$, three numerical parameters, namely, Π_0, Π_1, and Π_2, are used, which have the dimension of the surface tension coefficient and characterize the local properties of the isotherm near the equilibrium state of the free surface. Such a replacement has become possible because we used the power series of the surface tension coefficient in the vicinity of the equilibrium state of the free surface. As was noted above, Π_0 has the meaning of a surface tension coefficient on the planar (equilibrium) surface that is covered by a surfactant film with surface concentration Γ_0. Parameter Π_1 equals the surfactant concentration times the slope of the isotherm at point $\Gamma = \Gamma_0$. It is called the film elasticity. For conventional (not inactive) surfactants, $\Pi_1 < 0$. This parameter defines the force per unit length acting between two linear elements on the surface that have different surfactant concentrations. This force arises when the surfactant is nonuniformly distributed over the film and is directed along its surface. Parameter Π_2 depends on the curvature of the isotherm at point $\Gamma = \Gamma_0$.

Second, the new set involves parameter W, which characterizes the stability of the uniformly charged planar surface of the fluid against self-charge [14]. From the solution to the first-order problem [8, 10], it follows that

$$\operatorname{Re}(S) = r > 0, \quad \operatorname{Im}(S) = \omega = 0, \quad (5)$$

if

$$\omega_0^2 < 0, \quad \text{or} \quad W > \frac{1}{ak} + ak,$$

since complex frequency S in (4) takes the form $S = \pm i\omega_0$ in going to a perfect fluid in the absence of the surfactant (ν, D, Π_1) $\longrightarrow 0$. In this case, electric forces at the ridges of the waves with wavenumber k dominate over Laplace forces even in the first order of smallness. The surface becomes unstable against however small periodic wave perturbations for which relationship (5) is fulfilled; in other words, the charged surface of the fluid becomes unstable against self-charge [8, 10, 14]. From (5), it readily follows that all wavenumbers $k > 0$ are stable if $0 \le W < 2$. At $W = 2$, there appears wavenumber $k_* = 1/a$ lying at the boundary of instability in the sense that any however small increase of W above $W = 2$ makes the wave perturbation with $k = k_*$ unsta-

ble. At $W > 2$, there exists an interval of unstable wave-numbers that includes k_* and extends with increasing W.

If $r > 0$ and $t \longrightarrow \infty$ in expression (1), the ratio of the correction proportional to η^2 to the fundamental term proportional to η takes the form $\infty \times \eta$. This means that asymptotic expression (1) becomes nonuniform at large time intervals. Moreover, with condition (5) satisfied, the wave motion of the free surface ceases, since $\omega = 0$.

With regard to the aforesaid, we will study the profile of the wave under the assumption that the condition

$$W < \frac{1}{ak} + ak: \ \Rightarrow r = \mathrm{Re}(S) < 0, \quad \mathrm{Im}(S) = \omega \neq 0 \ (6)$$

is met.

In this case, parameter r has the meaning of the damping decrement of the wave in the first order of smallness and the correction-to-fundamental term ratio in expression (1) for the wave profile tends to zero in the limit $\eta \longrightarrow 0$ for any $t > 0$.

It is noteworthy that expression (1) for the profile can be alternatively written as

$$\xi = \eta \cos\theta \exp(rt) + \eta^2 A \cos(2\theta + \phi)\exp(2rt);$$

$$A = 2\sqrt{\mathrm{Re}(\zeta)^2 + \mathrm{Im}(\zeta)^2};$$

$$\phi = \begin{cases} \arctan\left(\dfrac{\mathrm{Im}(\zeta)}{\mathrm{Re}(\zeta)}\right), & \text{if} \quad \mathrm{Re}(\zeta) > 0; \\[2mm] \dfrac{\pi}{2}, & \text{if} \quad \mathrm{Re}(\zeta) = 0; \quad (7) \\[2mm] \arctan\left(\dfrac{\mathrm{Im}(\zeta)}{\mathrm{Re}(\zeta)}\right) + \pi, & \text{if} \quad \mathrm{Re}(\zeta) < 0. \end{cases}$$

Parameters ζ, r, and θ are the same as in (1).

Below, we will pass to the dimensionless variables where $\rho = g = \gamma = 1$ and the remaining quantities are expressed in terms of their characteristic scales:

$$k_* = \frac{1}{a}; \quad \eta^* = a; \quad \zeta_* = \frac{1}{a};$$

$$\Pi_n^* = \gamma; \quad \nu^* = \sqrt{ga^3}; \quad D^* = \sqrt{ga^3}.$$

4. INTRINSIC NONLINEAR INTERACTION OF WAVES IN THE ABSENCE OF SURFACTANT

In [1–4], expression (1) for the profile of a periodic nonlinear traveling capillary–gravitational wave was studied in the simplest case, i.e., in the absence of a surfactant film on the free surface. It was shown that the nonlinear behavior of the wave shows up most vividly when

$$k_* = 1/\sqrt{2}, \qquad (8)$$

where k_* is the dimensionless wavenumber.
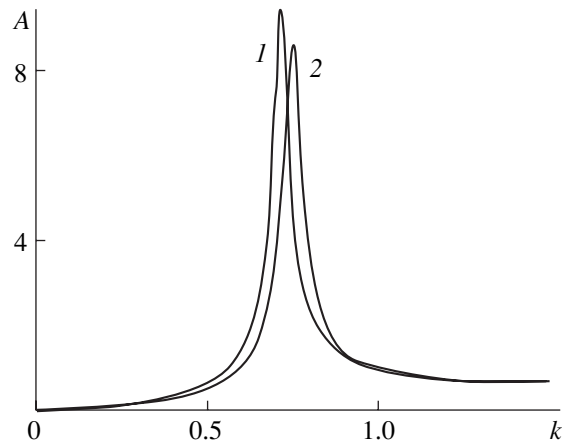
Fig. 1. Dimensionless second-order amplitude correction A to the wave profile vs. wavenumber k for $W = 0$ and $\Pi_1 =$ (1) 0 and (2) –0.4.

This follows from the resonance-like dependence $A = A(k)$ (Fig. 1). The height of the resonance peak characterizes the intensity of nonlinear interaction between the wave with wavenumber k (the fundamental wave term in (1), which is proportional to η; in the discussion which follows, it will be called the k wave) and the wave with wavenumber $2k$ (the correction proportional to η^2 or the $2k$ wave). The phase velocities of the waves are the same. A change in k has no effect on the amplitude η of the fundamental wave term but noticeably affect factor A (see (7)), which specifies the amplitude of the $2k$ wave. Thus, wavenumber k, which the wavelength of the fundamental wave depends on, affects the amplitude of the second-order correction. It seems as though the k wave and $2k$ wave interact with each other. In the publications concerned with the problems of nonlinear waves (see, e.g., [15]), such an interaction is called the intrinsic nonlinear interaction of waves. It is important that amplitude factor A in formula (7) serves as a measure of the interaction intensity.

The wave with wavenumber $2k$ is not an independent wave. Generally, its frequency and wavenumber do not satisfy the dispersion relation, unlike the fundamental (k) wave in (1). It satisfies the dispersion relation only at $k = k_*$. The phase velocity and amplitude of the $2k$ wave are completely defined by the k wave; in other words, the $2k$ wave is generated by the k wave and does not represent an independent wave motion.

The peak value of amplitude A and, hence, the intensity of nonlinear interaction depend on the viscosity of the fluid and surface charge. It was shown [1] that A monotonically grows with decreasing viscosity and goes to infinity in the limit of perfect fluid. This situation corresponds to the degenerate case of three-mode nonlinear resonance interaction between capillary and gravitational waves [12, 13]. If the viscosity is other than zero, the resonance peak height is finite [1–4].
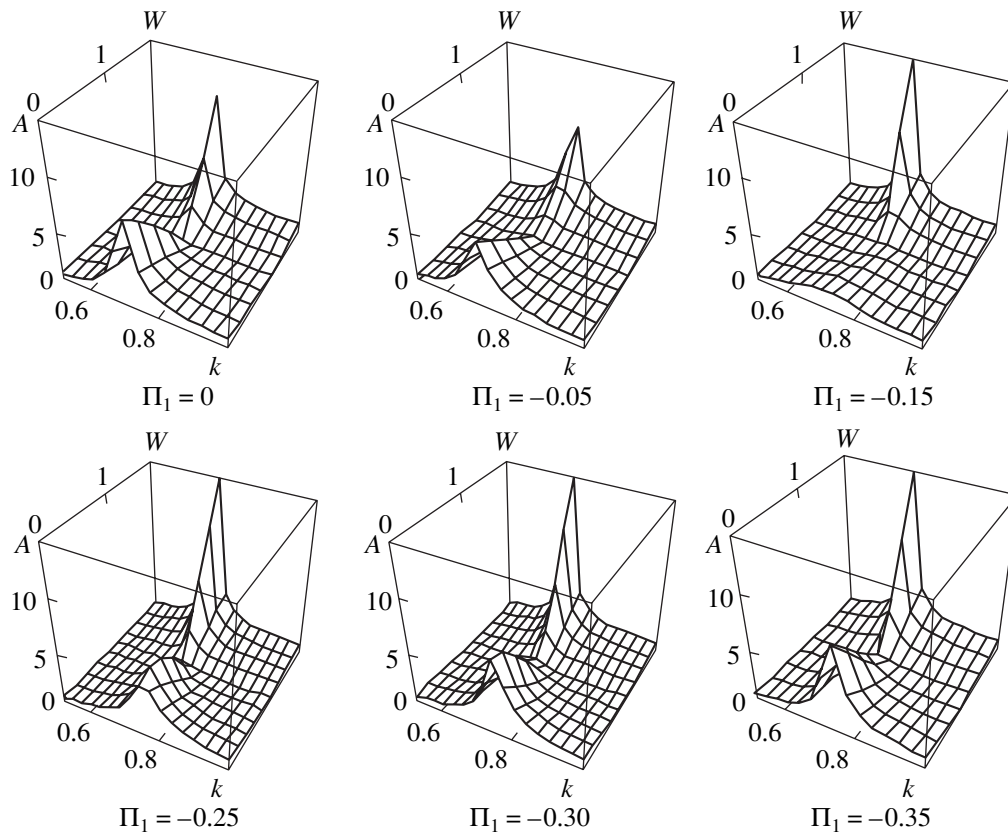
Fig. 2. Dimensionless second-order amplitude correction A to the wave profile vs. wavenumber k and surface charge density parameter W for different values of coefficient Π_1, which characterizes the elasticity of the surfactant film.

In the general case, a nonlinear periodic capillary–gravitational wave includes, along with the fundamental k wave and $2k$ correction, an infinite set of $3k$, $4k$, $5k$, etc., harmonics, which may interact with each other. We restrict our analysis to the simplest case of interaction.

It was noted [2] that the interaction intensity is a complex function of the surface charge density, the square of which is proportional to parameter W. It is known [3, 4] that, in the line

$$W = (k + k^{-1})/2, \qquad (9)$$

lying in the plane (k, W), the amplitude of the second-order term in (1) has a minimum tending to zero as the viscosity decreases. This means that, with such values of k and W, the wave motion of a perfect fluid does not contain the $2k$ wave; that is, the mechanism of excitation of this wave and, hence, the mechanism of action on it, fail in this situation. Then, solution (1) includes only the k wave, and that part of the solution responsible for the $2k$ wave (with the amplitude depending only on the properties of the fundamental wave) vanishes. On the other hand [3], there always exists an independent $2k$ wave of amplitude unrelated to the fundamental k wave. This $2k$ wave is a solution to the homogeneous part of the second-order problem and propagates with its own phase velocity, which coincides with the veloc-

ity of the k wave only at $k = k_*$. The frequency and wavenumber of this independent $2k$ wave satisfy the dispersion relation. In (1) and (7), the term responsible for the independent $2k$ wave (unrelated to the fundamental term) is omitted, since only that part of the solution responsible for nonlinear interaction is of interest in the context of our study. When k and W satisfy (9), the amplitude of the independent $2k$ wave may be set arbitrarily even if the amplitude of the k wave is zero and the waves with wavenumbers k and $2k$ propagate independently. If k and W satisfy (9) in the case of a viscous fluid, the intensity of nonlinear interaction is very weak.

Figure 2 shows the dependences $A = A(k, W)$ for different values of parameter Π_1. Line 1 in Fig. 1 corresponds to the case when the plane $W = \text{const}$ cuts the surface at $\Pi_1 = 0$. In Fig. 2, a resonance ridge above the straight line $k = k_*$ that lies in the parameter plane (k, W) is distinctly seen at $\Pi_1 = 0$. The ridge has a noticeable dip. The position of the dip ($k = k_* \approx 0.70$, $W = 0.5(k_* + k_*^{-1}) \approx 1.06$ is the point of intersection of the straight line $k = k_*$ and curve (9) on the plane (k, W). From Fig. 2, it is evident that, if a family of curves similar to line 1 in Fig. 1 is constructed for different W, the height of the resonance ridge on them will vary non-
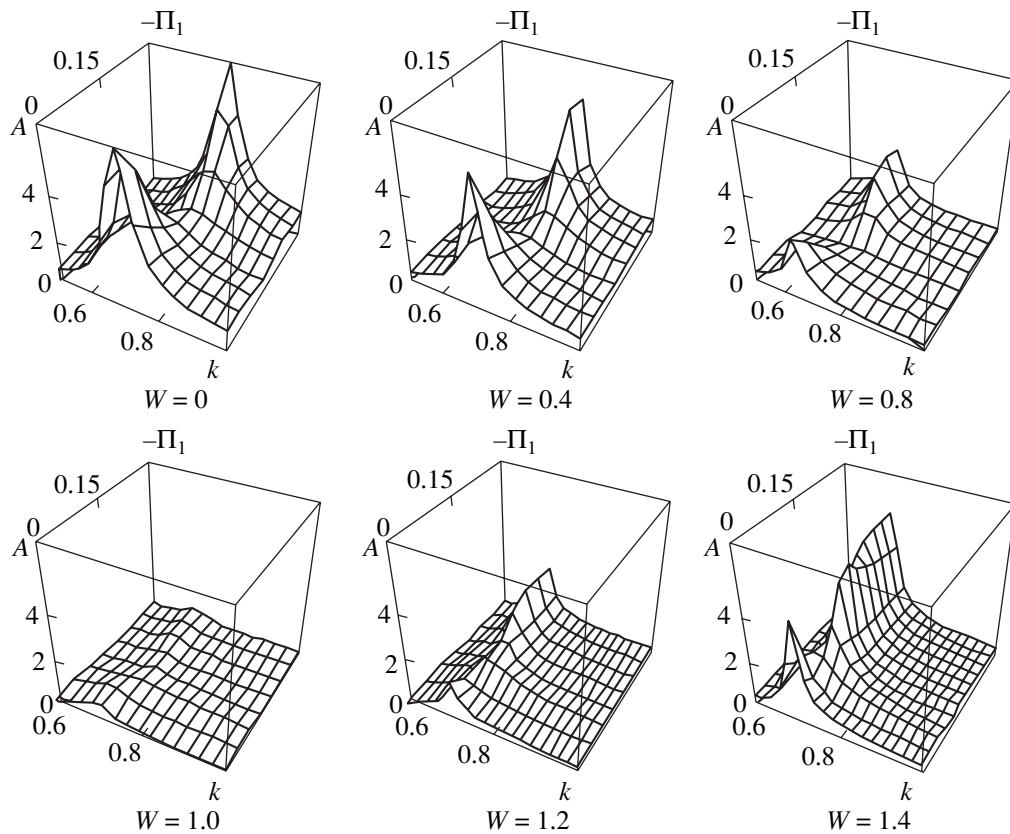
Fig. 3. Dimensionless second-order amplitude correction A to the wave profile vs. wavenumber k and dimensionless coefficient Π_1, which characterizes the elasticity of the surfactant film, for different values of surface charge density parameter W.

monotonically with increasing W. For W increasing from zero to 1.06, it decreases nearly to zero, since the position of the ridge in coordinates (k, W) approaches curve (9) of minimal nonlinear interaction. As W grows further, the ridge moves away from curve (9) and the resonance peak increases, reflecting an increase in the intensity of nonlinear interaction [3].

5. EFFECT OF SURFACTANT ON THE INTENSITY OF INTRINSIC NONLINEAR INTERACTION

To study the effect of a surfactant film on the intensity of nonlinear interaction, we will first clear up the dependence of amplitude factor A on various parameters.

The family of surfaces in Fig. 2 shows the dependence $A = A(k, W)$ at different values of film elasticity Π_1 (it is known [16] that surfactants on the surface of a liquid film diminish its surface tension, so that $\Pi_1 < 0$ in this case. Therefore, the illustrative calculations were performed for negative Π_1.). As follows from Fig. 2, the dependence $A = A(k, W)$ changes in a complex manner as the absolute value of Π_1, $|\Pi_1|$, increases. For a fixed value of dimensionless viscosity $\nu = 0.01$, the increase in $|\Pi_1|$ to ≈ 0.15 drastically reduces the interaction inten-

sity at small surface charge densities ($W < 1$). The resonance ridge in this dependence constructed for $|\Pi_1| \approx 0.15$ is hardly visible. A further increase in $|\Pi_1|$ enhances the interaction intensity. For $|\Pi_1| \approx 0.4$, we get nearly the same interaction pattern as for $|\Pi_1| = 0$. A closer look at these dependences shows that the elasticity of the surfactant film has an effect not only on the interaction intensity but also on the resonance wavenumber in (7). After the surfactant elasticity passes the value $|\Pi_1| \approx 0.15$, a further rise in $|\Pi_1|$ brings about a resonance ridge that originates (at $W = 0$) above the point with $k = k_* \approx 0.8$ rather than with $k = k_* \approx 0.7$. This is seen most distinctly in Fig. 1, where the dependences $A = A(k)$ are shown for $\Pi_1 = 0$ (curve 1) and $\Pi_1 = -0.4$ (curve 2). The values of the other parameters are the same as in Fig. 2. The shift of the resonance wavenumber depends on the viscosity and parameter W. For $\nu = 0.05$, $\Pi_1 = -0.4$, and $W = 0$, the resonance value of k is close to unity but the height of the resonance peak falls to unity, which is much lower than the heights of the peaks depicted in Fig. 1. An increase in W returns the resonance value of the wavenumber to k_* but now at $W \approx 1$. As W grows further, the position of the resonance ridge over the parameter plane (k, W) remains unchanged.

deep at $-\Pi_1 \approx 0.18$. Here, this is the value of the film elasticity at which the interaction intensity is minimal. The increase in W from zero to unity markedly decreases the height of the resonance ridge throughout its length, and its contours in Fig. 3 are smeared especially at small $-\Pi_1$. For $W > 1$, the ridge becomes sharply defined again. In all the curves shown in Fig. 3, the line of the resonance bridge originates above the point $k = k_*$. Then, the line of the ridge runs over the straight line $k = k_*$ lying in the plane $(k, -\Pi_1)$ until the ridge reaches a minimum. Thereafter, the ridge deviates slightly toward higher wavenumbers, $k > k_*$. This deviation is generally insignificant. It becomes appreciable only for $W > 1$ and increases with increasing viscosity ν.

Our analysis shows that, like surface charge density W, film elasticity Π_1 substantially influences the intensity of nonlinear interaction especially at near-resonance k. To illustrate this, we plotted A against W and $-\Pi_1$ for $k = k_*$ (Fig. 4). From Fig. 4, it follows that the dependence $A = A(-\Pi_1, W)$ at $k = k_*$ is very complicated. Moreover, the position of the folds and local extrema on this surface varies strongly with the viscosity. Perhaps the most interesting feature of this surface is that the least values of A are observed above the straight line $W = 1$, in the vicinity of which the interaction is minimal. However, these least values are other than zero and considerably depend on the elasticity of the surfactant film. Figure 5 demonstrates the curves that are obtained when the surface shown in Fig. 4 is cut by the planes $W = 0.8$, $W = 1.0$, and $W = 1.2$. For the closely spaced values of W, the difference in the effect of the surfactant on the intensity of nonlinear interaction is fairly pronounced. This means that, in the experiments, the effect of surfactants on the shape of nonlinear waves with wavenumber $k \approx k_*$ is expected to be very sensitive to the surface charge.

Fig. 4. Dependence $A = A(-\Pi_1, W)$ for $k = 1/\sqrt{2}$ and $\Pi_2 = D = 0$.

Fig. 5. Dependence $A = A(-\Pi_1)$ for $k = 1/\sqrt{2}$ and $\Pi_2 = D = 0$. $W = (1)$ 1, (2) 1.2, and (3) 0.8.

Figure 3 shows a family of curves $A = A(k, -\Pi_1)$ for different W. As in Fig. 2, cutting of the surface shown in Fig. 3 by the plane $\Pi_1 = 0$ gives line 1 in Fig. 1. Here, the second variable in the argument plane is $-\Pi_1$ instead of W as in Fig. 2. In going from the family of curves in Fig. 2 to that depicted in Fig. 3, parameters W and Π_1 exchange places. However, the curves in Figs. 2 and 3 qualitatively behave in a similar way. At $W = 0$, when only the elasticity of the film affects the intensity of nonlinear interaction between the waves, the resonance ridge is distinctly seen above the straight line $k = k_*$ lying in the parameter plane $(k, -\Pi_1)$. The ridge has a

CONCLUSIONS

A surfactant film considerably influences the profiles of nonlinear periodic capillary–gravitational waves with near-resonance wavenumbers (the doubled square of the resonance value equals unity divided by the capillary constant squared). The dependence of the intensity of nonlinear interaction between the harmonics constituting a nonlinear capillary–gravitational wave on the film elasticity is of nonmonotonic character. There exists a value of the elasticity that minimizes the interaction intensity. This value considerably depends on the viscosity of the fluid. The presence of a surfactant film increases the resonance wavenumber at which the interaction between the waves is the highest. The effect of a surfactant film on the intensity of nonlinear interaction depends on the surface charge in a complicated manner.

APPENDIX

Auxiliary Quantities and Relationships

(1) The unit vectors tangent and normal to the disturbed free surface are given by

$$\mathbf{n} = -\frac{\partial_x \xi}{\sqrt{1 + (\partial_x \xi)^2}} \mathbf{e}_x + \frac{1}{\sqrt{1 + (\partial_x \xi)^2}} \mathbf{e}_z;$$

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

(2) The right-hand sides of the expressions that are involved in the mathematical statement of the problem in the second order of smallness are

$$\mathbf{V}_2 = -\frac{1}{2}\mathbf{V}(\mathbf{U}_1^2) + \mathbf{U}_1 \times (\mathbf{V} \times \mathbf{U}_1);$$

$$f_{12} = \xi_1 \partial_z v_1 - u_1 \partial_x \xi_1;$$

$$f_{22} = 2\rho v \xi_1 \partial_{zz} v_1 - \xi_1 \partial_z p_1 - \frac{(\mathbf{V}\Phi_1)^2}{8\pi} + \frac{E_0}{4\pi}\xi_1 \partial_{zz}\Phi_1$$
$$- \chi(\Gamma_1 \partial_{xx} \xi_1 + 2\partial_x \xi_1 \partial_x \Gamma_1);$$

$$f_{32} = \rho v(4\partial_z v_1 \partial_x \xi_1 + \xi_1 \partial_z(\partial_z u_1 + \partial_x v_1) - \beta\Gamma_1 \partial_x \Gamma_1);$$

$$f_{42} = -\xi_1 \partial_z \Phi_1;$$

$$f_{52} = -\partial_x(u_1 \Gamma_1) - \Gamma_0\left(\xi_1 \partial_{xz} u_1 + \frac{\chi}{\rho v}\partial_x \Gamma_1 \partial_x \xi_1\right).$$

(3) Coefficients M_j have the form

$$M_j = \det \begin{pmatrix} 0 & -k & ik & 0 & R_{1j} \\ \dfrac{kE_0}{2\pi} & -\rho(S + 4vk^2) & 2\rho vik\sqrt{2\left(2k^2 + \dfrac{S}{v}\right)} & 0 & R_{2j} \\ 0 & 4i\rho vk^2 & \rho v\left(4k^2 + \dfrac{S}{v}\right) & -ik\Pi_1 & R_{3j} \\ 1 & 0 & 0 & 0 & R_{4j} \\ 0 & -2k^2\Pi_1 & ik\Pi_1\sqrt{2\left(2k^2 + \dfrac{S}{v}\right)} & (S + 2k^2 D)\Pi_1 & R_{5j} \end{pmatrix};$$

where

$$R_{10} = 2S; \quad R_{20} = -(\rho g + 4\gamma k^2); \quad R_{30} = 0;$$

$$R_{40} = -E_0; \quad R_{50} = 0;$$

$$R_{11} = \frac{1}{2}k$$

$$\times bk\left(1 - \frac{ic(k^2 - q^2)}{(3k + q)(2S + v(k - q)(3k + q))} - icq\right);$$

$$R_{21} = \frac{1}{4}k\left(bS\rho + 2\rho v(bk^2 - icq^2) + \frac{kE_0^2}{4\pi} + 3dk\Pi_1\right.$$
$$\left. - \frac{2\rho ibck(k - q)(2S + v(5k^2 + 2kq + q^2))}{(3k + q)(2S + v(k - q)(3k + q))}\right);$$

$$R_{41} = k\frac{E_0}{4\pi};$$

$$R_{31} = -\frac{1}{4}\left(\rho v(cq(5k^2 + q^2) + ik(6bk^2 - d^2\Pi_2))\right.$$
$$\left. + \frac{\rho vbck(k^2 - q^2)(5k^2 + 2kq + q^2)}{(3k + q)(2S + v(k - q)(3k + q))}\right);$$

$$R_{51} = \frac{1}{4}k\Pi_1\left(\frac{dk\Pi_1}{\rho v} - icq(2d + q)\right.$$
$$\left. + bk\left(2d + k - \frac{2ic(k - q)(k + q)^2}{(3k + q)(2S + v(k - q)(3k + q))}\right)\right);$$

$$b = -\frac{\omega_0^2 + 2vkqS}{k(S + 2vk(k - q))}; \quad c = \frac{i(\omega_0^2 + S(S + 2vk^2))}{k(S + 2v\kappa(k - q))};$$

$$d = \frac{\omega_0^2(q - k) + qS^2}{(S + Dk^2)(S + 2vk(q - k))};$$

$$q = \sqrt{k^2 + \frac{S}{v}}; \quad E_0 = 2\sqrt{\pi W \sqrt{\rho g \gamma}},$$

and i is the imaginary unit.

(4) The dimensionless dispersion relation has the form

$$\begin{cases} F(\alpha, \beta, \delta, \Lambda) = \sqrt{\alpha + \beta^4} \\ \mathrm{Re}(F(\alpha, \beta, \delta, \Lambda)) > 0, \end{cases}$$

where

$$\beta = \sqrt{\frac{\nu k^2}{\omega_0}}; \quad \Lambda = \frac{k^3 \Pi_1}{\rho \omega_0^2}; \quad \delta = \frac{D k^2}{\omega_0};$$

$$F(\alpha, \beta, \delta, \Lambda) = \left((\alpha + 2\beta^2)^2 + \left(1 + \frac{\Lambda}{\alpha(\alpha + \delta)} \right) \right)$$

$$\times \left(4\beta^2 \left(1 + \frac{\alpha^2 + 1}{\alpha(\alpha + \delta)} \frac{\Delta}{4\beta^4} \right) \right)^{-1}.$$

In general, this relation has two pairs of complex conjugate roots. Without loss in generality, the roots with negative imaginary parts may be omitted. This means that we select the roots responsible for the waves propagating in the positive direction of the *Ox* axis. Among the two other roots, one corresponds to a capillary–gravitational wave and the other to a wave associated with the presence of a surfactant film.

Let the two roots with positive imaginary parts be found for given $\beta = \beta_*$, $\Lambda = \Lambda_*$, and $\delta = \delta_*$. We put $\beta = \beta_*$, $\Lambda = 0$, and $\delta = \delta_*$ in the dimensional dispersion relation and will seek its root with a positive imaginary part. Then, by continuously varying Λ from zero to $\Lambda = \Lambda_*$, we will trace how the root found varies (this procedure is carried out numerically). For $\Lambda = \Lambda_*$, this root will become equal to one of those found initially and will be taken as corresponding to the capillary–gravitational wave.

ACKNOWLEDGMENTS

REFERENCES

1. D. F. Belonozhko and A. I. Grigor'ev, Zh. Tekh. Fiz. **73** (4), 28 (2003) [Tech. Phys. **48**, 404 (2003)].
2. D. F. Belonozhko and A. I. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **29** (8), 1 (2003) [Tech. Phys. Lett. **29**, 309 (2003)].
3. D. F. Belonozhko and A. I. Grigor'ev, Zh. Tekh. Fiz. **73** (11), 37 (2003) [Tech. Phys. **48**, 1396 (2003)].
4. D. F. Belonozhko and A. I. Grigor'ev, Zh. Tekh. Fiz. **74** (3), 5 (2004) [Tech. Phys. **49**, 287 (2004)].
5. V. G. Levich, *Physicochemical Hydrodynamics* (Fizmatgiz, Moscow, 1959) [in Russian].
6. E. H. Lucassen and J. Lucassen, Adv. Colloid Interface Sci. **2**, 347 (1969).
7. W. Alpers, J. Geophys. Res. **94**, 6251 (1989).
8. D. F. Belonozhko, S. O. Shiryaeva, and A. I. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **22** (15), 61 (1996) [Tech. Phys. Lett. **22**, 626 (1996)].
9. D. F. Belonozhko, A. I. Grigor'ev, and S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **23** (6), 75 (1997) [Tech. Phys. Lett. **23**, 242 (1997)].
10. S. O. Shiryaeva, D. F. Belonozhko, and A. I. Grigor'ev, Zh. Tekh. Fiz. **68** (2), 22 (1998) [Tech. Phys. **43**, 151 (1998)].
11. A. G. Boev and N. N. Yasnitskaya, Izv. Akad. Nauk, Fiz. Atmos. Okeana **39**, 132 (2003).
12. W. F. Simmons, Proc. R. Soc. London, Ser. A **309**, 551 (1969).
13. A. H. Nayfeh, J. Fluid Mech. **48**, 385 (1971).
14. Ya. I. Frenkel', Zh. Éksp. Teor. Fiz. **6**, 348 (1936).
15. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics* (Nauka, Moscow, 1988); R. Z. Sagdeev, D. A. Usikov, and G. M. Zaslavsky, *Nonlinear Physics: from the Pendulum to Turbulence and Chaos* (Harwood Academic, New York, 1988).
16. *Surface-Active Substances*, Ed. by A. A. Abramzon and G. M. Gaevskiĭ (Khimiya, Leningrad, 1979) [in Russian].

Translated by V. Isaakyan

=========== **GAS DISCHARGES,** ===========
PLASMA

Electron Self-Organization in Electronic Devices

V. G. Usychenko

St. Petersburg State Technical University, St. Petersburg, 195251 Russia
e-mail: Usychenko@rphf.spbstu.ru
Received April 6, 2004

Abstract—The open systems, which contain a huge number of electrons supplied from outside along with the energy, are described by a functional that takes into account the Lagrangians of all particles and is called the integral Lagrangian. A nonextremal principle is formulated that postulates that the value of this functional decreases as the system approaches the steady state. The principle is extended to the systems occurring near thermodynamic equilibrium (where it is virtually equivalent to the principle of minimum energy dissipation), as well as to the nonlinear systems, including those in which the motion of particles is described by equations of classical mechanics. The applicability of the principle is demonstrated by the examples of a vacuum diode, magnetron diode, and Gunn diode. © *2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Self-organization of systems containing a large number of elements starts from the appearance of stable links between elements. To this end, the elements should "feel" each other and mutually react, for example, on the existing fields. The electron systems are very attractive objects for investigations of the mechanisms of self-organization because of the long-range action of the Coulomb forces through which the electrons and the related structures interact with each other. The strong field domains in a Gunn diode [1, 2], the solitary electronic waves in a magnetron diode [3–6], the regular space charge oscillations in Penning cells [7], and the traveling waves in a plasma [8] are the manifestations of the collective behavior of charged particles self-organized into stable macroscopic structures called autowaves [9] under the action of applied constant electric and magnetic fields.

The self-organized electronic objects occur, as a rule, at high voltage gradients for which the principle of local equilibrium is not obeyed. In order to analyze the processes in vacuum devices, equations of electron motion are used that contain no dissipative terms in an explicit form. The particle energy in these devices is dissipated at the system boundaries, that is, at the surface of bombarded electrodes. The above properties radically distinguish electronic devices (especially vacuum ones) from dissipative systems, which are considered usually in nonequilibrium thermodynamics and synergetics [10–15]. The study of the mechanisms of particles' self-organization in such essentially nonequilibrium electronic systems requires a new approach, which is developed in this paper.

1. THE PRINCIPLE OF INTEGRAL LAGRANGIAN MINIMIZATION

Systems without electron drag. Let us consider an open system representing an electron vacuum device to the cathode and anode of which a constant voltage is applied from an external source. The cathode supplies matter (electrons) to the system. Accelerated by the external field, the particles travel to the anode acquiring energy from outside, (i.e., from the power supply). There are so many electrons that the strength of the self-consistent electric field they generate is comparable with the applied field strength. The establishment of the steady state, which proceeds with participation of the self-consistent fields, will be called the self-organization of particles. The motion of each electron will be described by the following equation:

$$m\ddot{x}_{ji} = F_{ji} + \delta F_{ji},$$

where m is the mass of the particle, $\ddot{x}_{ji}$ is the component of acceleration, F_{ji} is the component of a regular force acting on the particular electron at a given moment of time (j is the number of a particle), and δF_{ji} is the fluctuation force that takes into account individual interactions of the given electron with other particles.

The force $\mathbf{F}_j$ contains a potential force related to the existence of the self-consistent electric field, which is calculated using Poisson's equation. Thus, the cooperative, that is, integral action on the electron of all charge particles occurring in the volume V of the device, is taken into account. The random force $\delta\mathbf{F}_j$, which is small and only weakly affects the electron motion, is ignored in vacuum electronics [16]. The particles avoid collisions due to the long-range forces of mutual repulsion. A prominent example is offered by the motion of two opposite electron flows in a reflex klystron. For this

reason, a statistical description of the electron processes is based on the Vlasov equation, whereas at higher fields, the particles' motion equations are usually written in terms of the Lagrangean variables [16].

In a self-organizing system consisting of a large number of particles, the individual interactions cannot be ignored completely, since just these interactions play a leading role in the establishment of short-range correlations and in the formation of a microstructure of the collective formations [13]. However, since the role of short-range interactions (due to inequality $|\delta \mathbf{F}_j| \ll |\mathbf{F}_j|$) is correcting, rather than organizing, the electron trajectories, which should be corrected, are described without allowance for the random forces. This means that, in the course of particle self-organization at the first stages, we can use the following simplified equations:

$$m\ddot{x}_{ji} = F_{ji}. \tag{1}$$

The forces of individual interactions $\delta \mathbf{F}_j$ should be taken into account at the latest stages of the establishment of the steady state (this will be demonstrated below in the case of a magnetron diode). Thus, the behavior of an electronic device is described by a system of equations including Eq. (1) of particle motion, Poisson's equation, and the continuity equation.

Owing to the limited size of the device and the constant voltage $U_a > 0$ applied to the anode, the motion of electrons (at least of a considerable part of them), although described by time-reversible equations (1), is in fact irreversible: a jth electron emitted from the cathode at the moment t_{1j}, leaves the device forever after a lapse of time $\tau_j = t_{2j} - t_{1j}$, where t_{2j} is the time of electron impact on the anode.

Near the steady regime, the self-consistent field can be considered as quasi-static and the number of particles $N(t)$ counted at any instant of time $t \in \langle \tau \rangle$ as constant in the coordinate system moving synchronously with the autowave. Here, $\langle \tau \rangle = \sum_j \tau_j / N$ is the average particle lifetime in the device. The velocity vector of each electron and the value of the potential energy W_{pj} at each point of its path are single-valued functions of the coordinates of this point. Under these conditions, the Lagrangian for each particle $L_j(x_{ij}, \dot{x}_{ij}) = \sum_i 0.5\, m\dot{x}_{ij}^2 - W_{pj}$ does not depend explicitly on time. Let us sum the Lagrangians of all particles occurring in the volume V of the device and introduce the functional

$$\Lambda(t) = \sum_j L_j(t)/N(t) = \int_V L(t)n(t)dV / \int_V n(t)dV, \tag{2}$$

which will be called below the integral Lagrangian of the system. Here, $n(t)$ is the particle number density in the elemental volume dV and $L(t)$ is the sum of Lagrangians of all particles occurring in this volume. The values $n(t)$, $L(t)$, and $N(t)$ are taken at the same

instant of time $t \in \langle \tau \rangle$. In the steady regime, $N(t)$ and $\Lambda(t)$ take the stationary values N and Λ, respectively.

The experience gained in work with electronic devices suggests that, after a perturbation, the system always comes to the same steady state and does this regularly, rather than accidentally. Therefore, this specific steady state is energetically preferred to the other states. The energy parameters of the system are characterized by the functional Λ. Let us consider qualitatively the dependence of this quantity on the number N of particles.

At a given anode voltage $U_a > 0$ and near-vacuum (low pressure) conditions, the potentials in the interelectrode space and the corresponding particle velocities reach the highest values, for which the Lagrangian Λ approaches its upper limit. As the number of electrons increases, the value of Λ can only decrease, that is, in real processes we have $\partial \Lambda / \partial N < 0$. Indeed, an increase in the number N of particles possessing a negative charge is accompanied by their accumulation in a low-potential region, where the particle velocities are small and the residence time is long. Therefore, it would be reasonable to assume that, when the number of electrons N increases, the system always reorganizes in such a way that the integral Lagrangian reaches its minimal (as a rule, nonextremal) value

$$\Lambda_{\min} = (w_k - w_p)_{\min}. \tag{3}$$

In this expression, $w_k = W_k/N = \int_V W_{kV} n dV / \int_V n\, dV$ and $w_p = W_p/N = \int_V W_{pV} n dV / \int_V n\, dV$ take into account all forms of the kinetic and potential energy in the system; W_{kV} and W_{pV} are the kinetic energy of the particles and the potential energy of the system, respectively, in the volume element dV. The steady-state regime is established when, at given flows of energy and matter which connect the system with the environment, a further decrease of Lagrangian Λ becomes impossible. The evolution of the excited state is described by the expression

$$d\Lambda(t)/dt \le 0, \tag{4}$$

where the sign of equality is fulfilled as soon as value (3) is reached.

In the moving coordinate system, the time-independent integral Hamiltonian $H = $ const, representing the conservation of energy in the device space, is also a function of the steady state. In simple systems, the Hamiltonian $H = (w_k + w_p)$ contains the same functions w_k and w_p as those entering into the Lagrangian Λ. In such systems, to within insignificant constant values, condition (3) takes the form

$$\Lambda_{\min} = (w_k)_{\min} = (-w_p)_{\min}. \tag{5}$$

The obtained result discloses the teleological meaning of the nonextremal principle of the integral Lagrangian minimization (reduction): the system of

material particles always tends to a steady state with minimum potential energy. Since the work $A = -w_p$ should be performed in order to form the system of particles, an alternative formulation is also valid: for the given boundary conditions, a structure is realized, which requires minimum work to be spent for its formation.

Lagrangian Λ is a scalar. The minimum of the scalar value does not depend on the choice of the coordinate system [17]; therefore, the coordinates x_{ij} of the individual Lagrangians $L_j(x_{ij}, \dot{x}_{ij})$ can be considered as the generalized coordinates.

Systems with strong electron drag. Let us consider another limiting case, when the electron travels in the field $\mathbf{E}(x)$ in a space containing a large number of neutral atoms. Assuming that the scattering event is very short as compared to the average time τ_c between collisions, we can write the equation for the vector $\mathbf{u}$ of average velocity [18] as

$$m\dot{\mathbf{u}} = e\mathbf{E} - m\mathbf{u}\nu,$$

where e is the electron charge and ν is the effective frequency of collisions. Integrating this equation, we obtain the value $\mathbf{u}_d = e\mathbf{E}/m\nu$ for the average velocity, which is defined as a drift velocity of the particle [18].

Qualitatively the electron motion can be explained as follows [18]. Immediately after every effective collision, the electron moves in a random direction; therefore, the vector of average velocity is $\mathbf{u} = 0$. Up to the next collision, the electron moving in the field gains a directional velocity u_d and its kinetic energy increases by the value $eEu_d/\nu = mu_d^2 = -eU_d$, where $U_d = -\int_0^d E(x)dx$ is the potential difference across the electron mean free path d in the direction parallel to the electric field vector. Upon the next collision, this new portion of energy is dissipated as well, increasing the electron temperature in the system. Between collisions, the electron moves in the accelerating field and its averaged Lagrangian has the form $L = mu_\xi^2 - eU_\xi + 0.5mu_{ch}^2$. Here, $0.5mu_{ch}^2$ is the energy of chaotic motion and $mu_\xi^2 = -eU_\xi$ is the mean energy gained by the particle on the passage through the potential difference $U_\xi = -\int_0^\xi E(x)dx$, where $0 \le \xi \le d$. The inclusion of all N particles present in the system leads to the Lagrangian

$$\Lambda = 2m\langle u_\xi^2 \rangle + 0.5m\langle u_{ch}^2 \rangle.$$

Expressing the kinetic energy of the directional electron motion as $m\langle u_\xi^2 \rangle = \alpha mu_d^2$, where the constant $\alpha < 1$, and assuming that the mean chaos energy $0.5m\langle u_{ch}^2 \rangle$ is stationary, we obtain (to within insignifi-

cant constant values)

$$\Lambda_{min} = m(u_d^2)_{min}. \tag{6}$$

This formula has a simple physical meaning: without distant interaction with a neutral particle, the electron cannot avoid collision with this particle. Under such conditions, minimization of the Lagrangian is possible only through a decrease in the drift velocity. Equation (6) is also applicable as the first approximation to solid-state devices in a weak field, where the primary mechanism of energy dissipation is the scattering of carriers on acoustic phonons. In this case, m is the effective mass of charge carriers.

The essentially heuristic principle of Lagrangian minimization needs verification. Since Λ in the steady regime reaches a steady rather than extremal value, the usual variational methods are inapplicable and experimental data should be invoked. To this end, we will turn to well-known electronic devices whose theory perfectly fits to the experiment, namely, the vacuum diode (VD), magnetron diode (MD), and Gunn diode (GD). We will check the accuracy of realization of the principle of Λ minimization in various steady states, simultaneously taking into account the possible mechanisms of the self-organization of particles. The VD is interesting as a system in which all possible states occur on the thermodynamic branch. The VD approaches to thermodynamically equilibrium state when this brunch originates through a decrease in the emission and the anodic voltage $U_a = 0$. The MD is also of interest because its steady states are separated from the thermodynamic brunch by the point of instability. The GD is interesting as a system whose states can occur both on the thermodynamic branch and beyond, similarly to the case of an MD. Moreover, unlike the VD and MD, a GD is a solid-state rather than vacuum device.

2. VACUUM DIODE

Lagrangian. Let us introduce the cylindrical system of coordinates (r, φ, z) and consider a VD in which the distance between the anode and cathode is small as compared to the longitudinal size l of the diode along the z axis. The Lagrangian for an individual electron leaving the cathode with thermal velocity u_t has the form

$$L = 0.5m\dot{r}^2 - eU(r) + 0.5mu_t^2, \tag{7}$$

where $U(r)$ is the potential at the point where the particle is situated.

The kinetic energy

$$0.5m\dot{r}^2(r) = -eU(r) \tag{8}$$

acquired by the electron under the influence of the potential determines the radial velocity $\dot{r}(r)$. In the case of strong and uniform emission, azimuthally symmetric electron cloud is formed, which is characterized by a steady distribution of the particle concentration $n(r)$

and the corresponding space charge potential $U_{sc}(r) \leq 0$. Let us represent the potential $U(r)$ in the diode as

$$U(r) = U_v(r) + U_{sc}(r), \qquad (9)$$

where

$$U_v(r) = U_a \frac{\ln r/r_c}{\ln r_a/r_c} \geq 0$$

is the potential distribution in vacuum.

Substituting (9) into (7) and summing Lagrangians of all N electrons occurring in the VD volume $V = \pi l(r_a^2 - r_c^2)$, we obtain (after normalization to N)

$$\Lambda = w_{kr} + (w_N - w_{sc}) + kT_c. \qquad (10)$$

In this expression, $w_N = W_N/N$, where $W_N = -2\pi l \int_{r_c}^{r_a} n(r)U_v(r)r dr > 0$ is the energy supplied to the system from outside. According to the mean value theorem, $w_N = -e\beta U_a$, where $0 < \beta < 1$. The value $w_{sc} = W_{sc}/N$, where $W_{sc} = 2\pi le \int_{r_c}^{r_a} n(r)U_{sc}(r)r dr > 0$, is the potential energy of the space charge. The term $w_{kr} = W_{kr}/N$, where $W_{kr} = \pi lm \int_{r_c}^{r_a} n(r)\dot{r}^2 r dr$, is the kinetic energy of electrons moving relative to the steady space charge. The term kT_c takes into account the electron thermal energy: T_c is the cathode temperature and k is the Boltzmann constant. We will ignore the thermal energy for the anode voltage obeying the condition $U_a \gg -kT_c/e$.

Steady state. In view of (8), Lagrangian (10) can be reduced to form (5): $\Lambda_{min} = 2(w_{kr})_{min} = 2(w_N - w_{sc})_{min}$. Thus, we obtain the steady state condition

$$(w_{kr})_{min} = (w_N - w_{sc})_{min}. \qquad (11)$$

It follows from Eqs. (8), (9), and (11) that the value w_{kr} will be the smaller the higher the fraction of particles (of their total number N) localized in the region of small values of velocity $\dot{r}(r)$, which is determined by the potential $U(r)$, is. Thus, all states of a VD are characterized by the tendency of particles to reach the rest state at the bottom of the potential well near the cath-

Fig. 1. Trajectories of (a) an individual electron in vacuum and (b) electrons in Brillouin's "bushing."

ode. Here, their concentration causes the most pronounced decrease of $w_N - w_{sc}$. Considering the structure of the expressions $w_N = W_N/N$, $w_{sc} = W_{sc}/N$, and $w_{kr} = W_{kr}/N$, one can see that, as the number of particles N increases especially rapidly (in a first approximation in proportion to N), w_{sc} increases, whereas w_N and w_{kr} depend weakly on N. Therefore, the minimum of Lagrangian (10) is reached as the following condition is fulfilled:

$$w_{sc} = (w_{sc})_{max}. \qquad (12)$$

3. MAGNETRON DIODE

Lagrangian. Let us place an MD into a homogeneous magnetic field B directed along the z axis. As B increases starting from zero, the electron distribution in the MD will be azimuthally uniform unless a critical value of the magnetic field B_{cr} is reached, at which the electron returns to the cathode after nearly touching the anode. The value $B = B_{cr}$ corresponds to a bifurcation point. We will be interested in the regimes with the values $B > B_{cr}$ for which the apex of the trajectory of an electron starting from the cathode into vacuum is closer to the cathode than to the anode (Fig. 1a).

On the basis of the additive properties of Λ, we can make preliminary conclusions on the behavior of the system of particles through consideration of the Lagrange function for an individual electron emitted from the cathode with zero initial velocity,

$$L = m/2(\dot{r}^2 + r^2\dot{\varphi}^2 + \dot{z}^2) - eU - m/2r^2\omega\dot{\varphi}. \qquad (13)$$

Here, $\omega = -eB/m$ is the cyclotron frequency and $\dot{\varphi}(r) = \omega/2(1 - r_c^2/r^2)$ is the azimuthal velocity of the particle (for an azimuthally-symmetrical field distribution $(\partial U/\partial \varphi = 0)$; this velocity is independent of the potential $U(r)$ [19]). With regard to Hamiltonian $H = m/2(\dot{r}^2 + r^2\dot{\varphi}^2 + \dot{z}^2) + eU = 0$, the Lagrangian takes the form $L = -2eU - m/4r^2\omega^2(1 - r_k^2/r^2)$, from which it follows that the L value decreases as r increases. Therefore, in the course of self-organization, the electrons will tend to reach the maximum value of $r = r_a$. Indeed, the experiment [4] showed that, starting from very small emission currents (five to six orders of magnitude lower than the nominal ones), stable electron structures (called solitary waves) are formed in an MD [6, 20], in which a regular motion around the cathode is characterized by stable oscillations and current flow to the anode.

Let us consider a regime with a single wave (Fig. 2), which propagates around the cathode at a constant angular velocity $\Omega \ll \omega$. In the coordinate system $(r, \psi = \varphi - \Omega t, z)$ rotating together with the wave,
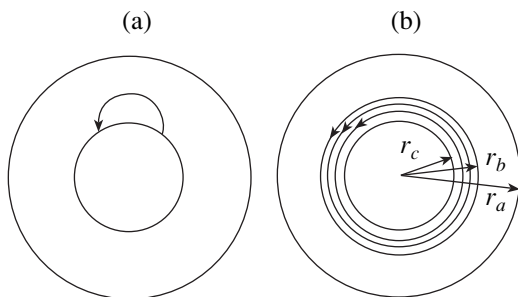
Lagrange function (13) takes the form [20]

$$L = 0.5m(\dot{r}^2 + r^2\dot{\psi}^2 + \dot{z}^2)$$

$$-[eU - 0.5mr^2\Omega(\Omega - \omega) - 0.5mr^2\dot{\psi}(2\Omega - \omega)],$$

where the terms in the square brackets represent various forms of the potential energy, including those related to the velocity-dependent forces. Summing the Lagrangians of all N electrons that occurred in the volume $V = \int_V dV = \int_{r_c}^{r_a} \int_0^{2\pi} \int_{-l/2}^{l/2} r\,dr\,d\psi\,dz$ and normalizing the sum to the total number of particles, we obtain the functional

$$\Lambda = w_{kr} + (w_N - w_{sc} + w_\Omega + w_\psi). \qquad (14)$$

Here, the term $w_{kr} = \dfrac{m}{2N}\int_V n\,(\dot{r}^2 + r^2\dot{\psi}^2 + \dot{z}^2)dV$ in the right-hand side of this equation represents the averaged kinetic energy of the relative motion of particles; the energies $w_\Omega = \dfrac{m}{2N}\int_V n\,r^2\Omega(\Omega - \omega)dV$ and $w_\psi = \dfrac{m}{2N}\int_V n\,r^2\dot{\psi}\,(2\Omega - \omega)dV$ are related to the generalized potentials determining the velocity-dependent forces; and $n = n(r, \psi, z)$ is the particle number density. The physical meaning of the rest terms is the same as in Eq. (10).

Self-organization. Comparing Eqs. (14) and (3), we obtain $w_k = w_{kr}$, $-w_p = w_N - w_{sc} + w_\Omega + w_\psi$. Since both $w_{kr} > 0$ and $-w_p > 0$, the steady-state condition takes the form

$$\Lambda_{min} = (w_{kr})_{min} + (w_N - w_{sc} + w_\Omega + w_\psi)_{min}. \qquad (15)$$

Among the energies entering into this equation, the value of w_{sc} increases especially rapidly with the number of particles N. Therefore, in an MD, as well as in a VD, the minimum of Λ is reached as $w_{sc} = (w_{sc})_{max}$. This condition is fulfilled by the hypothesis of Brillouin [19], according to which the space charge will be accumulated near the cathode as the emission increases and the potential $U(r) = U_v(r) + U_{sc}(r)$ will be reduced unless the kinetic energy w_{kr} reaches its minimum value. In this case, an electron "bushing" is formed around the cathode, where particles moving along the circular trajectories (Fig. 1b) can stay for a long time. Computer simulations [21, 22] and experiment [4] (indirectly) give an indication of the layered (i.e., ordered) motion of the particles. The transition to such motion [23] under the action of collision forces $\delta\mathbf{F}_j$ minimizes the effect of these forces. The electron-electron collisions changing the particle momenta [18] would result in an increase of the size of electron "bushing" and, accordingly, in a decrease in the w_{sc} value, that is, in an increase of Λ, which is in conflict with the requirement that $\partial\Lambda/\partial N < 0$. During the layered motion, the particles do not collide and, hence, their concentra-
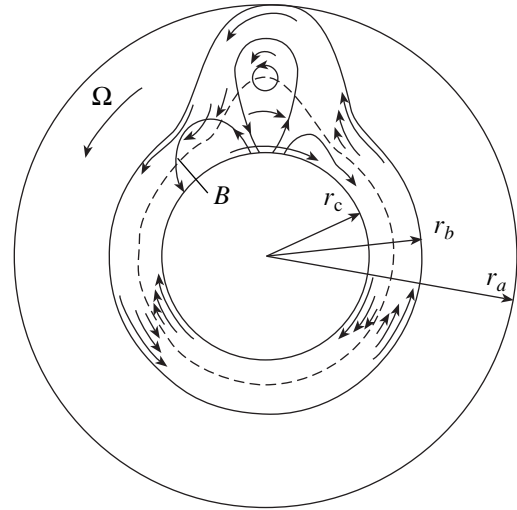
Fig. 2. Space charge structure in an MD at low electron emission in a rotating coordinate system. The lengths and directions of arrows characterize particle velocities. The dashed line denotes the separatrix, on which the velocities of electrons are equal to zero; B is the trajectory of the electron bombarding the cathode.

tion and energy w_{sc} can reach the maximum values, whereas the Lagrangian can reach extremely low values.

However, as was indicated above, the electrons in the MD should reach the anode, which is not evident from Brillouin's solution. It was established later [21, 16] that the electron "bushing" is unstable and the development of instability results in the appearance of solitary waves. Figure 2 shows the MD cross section in the low-emission mode [6], which is sufficient for the formation of only one wave. The appearance of this wave does not violate the condition of the w_{sc} maximum in the rest part of the "bushing." However, the electron number density in the wave is extremely high as well. The structure of solitary waves determined [6] based on this assumption showed that the dynamic parameters of these waves are in a good agreement with the results of measurements.

Evolution. The electrons emitted from the frontal part of the wave (Fig. 2) enter its azimuthal field and take a part of the wave energy that goes on the bombardment of the cathode. In formula (14), this energy enters into the term w_{kr}. The cathode temperature increases as a result of bombardment [24] and, hence, the wave controls in this way the cathode emission and its own development. Experiment [4] and theory [6] showed that the number of both waves and electrons in these waves are increased with the growing emission. Since, as N increases, w_{sc} changes faster than any other term in (14) (thus determining the variation of the Lagrangian as a whole), evolution principle (4) for a system capable of controlling the amount of supplied matter can be expressed in another (equivalent) way (in

the normalized and nonnormalized forms, respectively):

$$dw_{sc}/dt \geq 0, \quad dW_{sc}/dt \geq 0. \quad (16)$$

If no measures are provided in the MD to limit the cathode bombardment, then the continuous growth of the number of emitted electrons is accompanied by the increase on the potential energy W_{sc} of the waves, which will eventually result in their decay [6]. The MD passes to the chaotic state [4, 6, 25].

4. GUNN DIODE

In deriving formula (6), we assumed implicitly that the potential difference over the electron mean free path is known. This value can be readily determined, provided that we know the voltage–velocity characteristic of electrons and the voltage applied to a GD. However, we will use another way that unravels more comprehensively the synergetic content of the phenomena under consideration.

The formation of a high-field domain in a GD is related to the instability [2, 26] stemming from the transition of electrons accelerated by the external field from the bottom to upper valley, whereby a region with negative steepness appears in the dependence of electron drift velocity $u_d(E)$ on the electric field (see Fig. 3 [27]). If the intervalley transition did not take place, that is, the energy dissipation characteristics were the same as in the low field, then the drift velocity of the particles would continue to depend linearly on the field E, as is shown by the dash-and-dot line in Fig. 3. The total
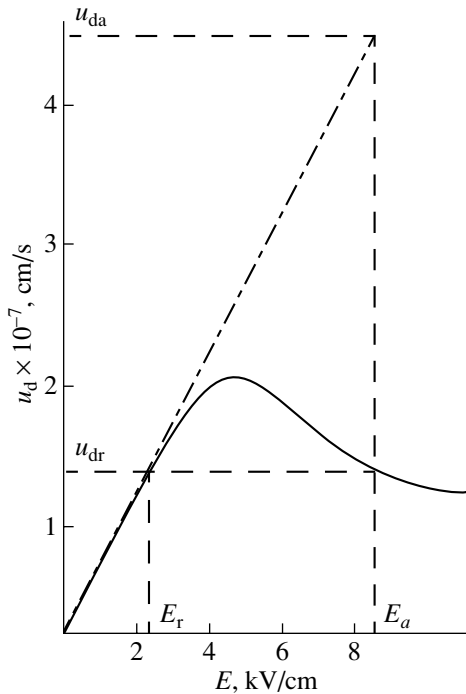
Fig. 3. Electron velocity in a GD as a function of the applied electric field.

energy W_N acquired by N electrons within a time τ_c between two collisions with the lattice would be equal to the mean drift kinetic energy:

$$W_N = n_0 s l m_1 u_{da}^2 = n_0 s l m_1 \mu_1^2 E_a^2 = -eN U_a \tau_c/\tau_a. \quad (17)$$

Here, l and s are the length of diode and its cross-section area, respectively; n_0 and μ_1 are the low-field electron concentration and mobility, respectively; $E_a = U_a/l$ is the absolute value of the average field in the diode; U_a is the anode voltage; and $\tau_a = l/u_{da}$ is the drift time of particles; $\tau_c = -(m_1\mu_1)/e$. The rest of the designations follow from Fig. 3. Considering that, on transition to the upper valley, a portion of the electron energy is spent for the domain formation without dissipation in the lattice and assuming the domain voltage to be $U_d = l(E_a - E_r)$, we will find the potential energy of the domain to be

$$W_{sc} = n_0 s l m_1 \mu_1^2 (E_a^2 - E_r^2) = W_N(1 - E_r^2/E_a^2). \quad (18)$$

The domain travels over the diode at the velocity $u_{dr} < u_{da}$. The total value of the drift kinetic energy of all particles in the diode is

$$W_k = W_N - W_{sc} = N m_1 u_{dr}^2 = W_N E_r^2/E_a^2. \quad (19)$$

This analysis of the domain mode in a GD is different from that described in the literature [2, 26] and thus should be verified. To this end, we compare the energy W_{sc} and the energy $0.5 C_d U_d^2$ stored in the equivalent capacity C_d of the domain and determine the domain capacity:

$$C_d = \frac{2 s n_0 m_1 \mu_1^2 (E_a + E_r)}{l(E_a - E_r)}.$$

Calculations using this formula, as well as formulas from [2, 26] and those obtained from other physical prerequisites, gave the values of C_d differing less than by a factor of 1.5. Such a correspondence can be considered as quite satisfactory.

Self-organization. The number N of electrons in a GD is constant, and the domain formation is only possible if these electrons are redistributed over the diode. Let us introduce the quantity N_1, which is the number of electrons in the domain. The kinetic energy

$$W_{ks} = W_N \frac{N_1 E_r^2}{N E_a^2}$$

of the collective structure (domain), as well as its total energy,

$$W_\Sigma = W_{sc} + W_{ks} = -eN U_a \frac{\tau_c}{\tau_a}\left[1 - \frac{E_r^2}{E_a^2}\left(1 - \frac{N_1}{N}\right)\right],$$

are dissipated neither during the drift nor in the course of domain "loss," since, as soon as one domain disap-

pears at the anode, another energetically equivalent domain is generated at the cathode [2]. Equation (17) takes the form

$$W_N = W_{sc} + W_{ks} + W_{kr}^{\downarrow}, \qquad (20)$$

where

$$W_{kr}^{\downarrow} = -eNU_a \frac{\tau_c E_r^2}{\tau_a E_a^2}(1 - N_1/N) \qquad (21)$$

is the drift energy of those $N - N_1$ electrons that occur outside the domain. Just this energy transforms into the thermal energy, as the electrons are scattered in the lattice. This energy also determines the value of functional (6), for which we obtain the following expression, on the basis of (5) and (20):

$$\begin{aligned} \Lambda_{min} &= m_1(u_{dr}^2)_{min} = (w_{kr}^{\downarrow})_{min} \\ &= (w_N - w_{sc} - w_{ks})_{min}. \end{aligned} \qquad (22)$$

The value of $w_{kr}^{\downarrow}$ decreases as the number of particles N_1 in the domain increases, that is, as w_{sc} grows. Thus, in a GD, as well as in a VD and MD, the minimum of Λ is reached when $w_{sc} = (w_{sc})_{max}$.

5. COMPARISON WITH THE PRINCIPLES OF NONEQUILIBRIUM THERMODYNAMICS

In the case of a small deviation from the thermodynamic equilibrium, when the Onsager reciprocity relations are valid, the steady states of the open system obeying the local equilibrium conditions [10, 11] are characterized by the extremal principle of minimal entropy production. Gyarmati [28] reduced this principle to the more general variational principle of the thermodynamics of irreversible processes, namely, to the principle of minimum energy dissipation, which, in turn, was empirically generalized by Moiseev [29]. The practical regimes of the systems under consideration are far from thermodynamic equilibrium. Therefore, in order to compare the principle of minimization of the Lagrangian Λ with the principle of minimum energy dissipation, we will treat the latter principle in the wording extended by Moiseev [29], according to whom "if a set of states conforming to the laws of conservation and relationships imposed onto the system is admissible, rather than a single state, then a state will be realized that corresponds to a minimum energy dissipation." We assume that this state is determined by the following condition:

$$(p_{dis})_{min} = (dw_{dis}/dt)_{min}, \qquad (23)$$

where $w_{dis} = W_{dis}/N$ is the dissipation energy, which was not found previously, since it does not enter explicitly into the Lagrangian.

In order to solve the problem, we will turn to the energy balance equations for each system and deter-

mine an energy related to the term w_{dis} and, hence, entering into the integral Lagrangian. The control parameter is the number of particles N.

Vacuum diode. The Hamiltonian for one electron leaving the cathode with the thermal velocity u_t has the form

$$H = 0.5m\dot{r}^2 + eU = 0.5mu_t^2.$$

Summing the Hamiltonians of all N electrons occurring in the VD volume and taking into account relation (9), we obtain the energy balance equation

$$W_N + NkT_c = W_{sc} + W_{kr}.$$

All values in this equation are determined in (10). Let the intensity of electron emission be such that each electron reaches the anode. N particles occurring in the interelectrode space and traveling to the cathode possess the total kinetic energy W_{kr} and carry the current I_a, thus dissipating the power U_aI_a on the anode. During their life time $\langle\tau\rangle = (r_a - r_c)/\langle\dot{r}\rangle$, where $\langle\dot{r}\rangle$ is the average particle velocity, N electrons will dissipate the energy

$$W_a = W_{dis} = U_aI_a\langle\tau\rangle = -eNU_a. \qquad (24)$$

Since W_a and W_{kr} are two forms of the kinetic energy of the same electrons, we can introduce the relationship

$$W_{kr} = \theta W_a, \qquad (25)$$

where $\theta = \langle\dot{r}^2\rangle/\dot{r}_a^2 \leq 1$ and $\dot{r}_a$ is the electron velocity at the moment of impact on the anode. The coefficient θ characterizes inhomogeneity of the spatial distribution of the kinetic energy of the particles. The above equations yield

$$W_{kr} = -\theta eNU_a = \theta W_a = \theta W_{dis}, \quad p_{kr} \cong \frac{w_{kr}}{\langle\tau\rangle} = \theta p_{dis},$$
$$\qquad (26)$$
$$p_{dis} \cong \frac{w_a}{\langle\tau\rangle} = \frac{w_{kr}}{\theta\langle\tau\rangle}.$$

Thus, the dissipated energy $w_{dis} = w_a$ in the Lagrangian of a VD is taken into account implicitly by the term w_{kr}. Let us find the quantitative relationships between w_{kr} and $w_{dis} = w_a$ in various regimes. In order to simplify the calculations, we will consider a quasi-flat VD, in which $(r_a - r_c)/r_c \ll 1$.

Close to the thermodynamic equilibrium, the state of the VD is characterized by low emission ($W_{sc} \cong 0$) and by the value $U_a \ll -KT_c/e$. The electrons mainly possess velocity u_t, and it follows from (25), (26) that $\theta = \langle\dot{r}^2\rangle/\dot{r}_a^2 = 1$, $w_{kr} = w_{dis} = kT_c$. We can see from these relations that, near the thermodynamic equilibrium, the principle of Lagrangian minimization and the principle of minimal dissipation of energy give the same results.

Let us follow the evolution of the steady state with increasing the number N of particles in the strongly nonequilibrium regimes determined by the values $U_a =$

const $\gg -kT_c/e$. For a regime with limited emission, when each electron reaches the anode, it is easy to obtain $\langle \dot{r}^2 \rangle = 0.5\dot{r}_a^2$, $\theta = 0.5$ and $(w_{kr})_{min} = 0.5(w_{dis})_{min}$, $(p_{kr})_{min} = 0.5(p_{dis})_{min}$. As the electron emission increases infinitely, a near-cathode potential minimum arises [30, 31]. The majority of electrons unable to overcome this region return to the cathode. In this case, w_{sc} increases, w_{kr} decreases (according to (11)), the anode current tends to saturation, and $\theta = w_{kr}/w_a$ monotonically decreases. Thus, in the entire thermodynamic branch, the principle of minimum energy dissipation and the principle of minimization of Lagrangian Λ involve proportional values, $(p_{dis})_{min} \propto (p_{kr})_{min}$, but since $(p_{kr})_{min} \leq (p_{dis})_{min}$, the principle of Lagrangian minimization poses more rigid constraints on the system. The principle of Λ minimization is more informative, since it not only explains the reasons for the decrease of w_{kr} and p_{kr} caused by the space charge increase, but also points to the diode area where the space charge is accumulated.

Magnetron diode. In the coordinate system rotating with the angular velocity Ω of the wave, the Hamiltonian of a single electron leaving the cathode with zero velocity has the form [20] (we neglect the thermal energy of the particles)

$$H = \frac{m}{2}(\dot{r}^2 + r^2\dot{\psi}^2 + \dot{z}^2) + \frac{m}{2}\omega\Omega r^2$$
$$-\frac{m}{2}r^2\Omega^2 + eU = \frac{m}{2}r_c^2\omega\Omega. \tag{27}$$

Summing the Hamiltonians of all particles in the MD volume, we obtain the equation of the integral energy balance,

$$W_N = W_{sc} + W_{kr} + W_{ks1} + W_{ks2}. \tag{28}$$

The terms $W_{ks1} = 0.5m\int_V n(r^2 - r_c^2)\omega\Omega dV$ and $W_{ks2} = -0.5m\int_V n r^2\Omega^2 dV$ in the right-hand side reflect the work performed by the collective structure (wave) to overcome back electromotive and centrifugal forces. Being kinetic in form, these energies are related to the generalized potentials by which the velocity-dependent forces are expressed. The physical meaning of other terms was determined above.

Using Hamiltonian (27) we can determine the energy dissipated on the anode during the lifetime $\langle \tau \rangle = -eN/I_a$ by all N electrons occurring in the MD volume:

$$W_a = 0.5m(\dot{r}_a^2 + r_a^2\dot{\psi}_a^2 + \dot{z}_a^2)N = -\zeta eNU_a. \tag{29}$$

Here $\dot{r}_a$, $r_a\dot{\psi}_a$, and $\dot{z}_a$ are the particle velocity components at the moment of impact on the anode;

$$\zeta = 1 + \frac{m}{2eU_a}(\omega\Omega(r_a^2 - r_c^2) - \Omega^2 r_a^2), \quad 0 < \zeta < 1 \tag{30}$$

is the coefficient [6] reflecting the fact that the energy $W_{ks1} + W_{ks2}$ of the collective forms of motion is not dissipated. Similarly to the case of a VD, the energy of the relative particle motion W_{kr} can be expressed in terms of the energy dissipated on the anode as

$$W_{kr} = -\theta\zeta eNU_a = \theta W_a = \theta W_{dis}.$$

Here, the coefficient

$$\theta = \frac{\langle \dot{r}^2 + r^2\dot{\psi}^2 + \dot{z}^2 \rangle}{\langle \dot{r}_a^2 + r_a^2\dot{\psi}_a^2 + \dot{z}_a^2 \rangle} < 1$$

characterizes the inhomogeneity of the distribution of the kinetic energy of the relative particle motion and the angular brackets denote the averaging over all electrons. Thus, the dissipated energy $w_{dis} = w_a$ is implicitly contained in Lagrangian (15) in the term w_{kr}. The requirement of a minimum of the value $w_{dis} = w_a$ imposed by the principle of minimum energy dissipation coincides with the condition of w_{kr} minimization in formula (15). However, since $w_{kr} < w_a$, the principle of Lagrangian minimization imposes more rigid constraints.

Gunn diode. The energy in a GD is dissipated only by $N - N_1$ electrons occurring outside the domain. The energy of these particles during their drift time $\tau_r = l/u_r$ is determined by the relationship

$$W_{kr} = W_{kr}^{\downarrow}\frac{\tau_r}{\tau_c} = -eNU_a\left(1 - \frac{N_1}{N}\right)\frac{E_r}{E_a}. \tag{31}$$

The particle velocity distribution is close to the Maxwell distribution; therefore, $\theta \cong 1$ and practically the entire energy W_{kr} converts into the thermal energy W_a. In the normalized form, the outlined effect reduces to the relations

$$w_{kr} = w_a = w_{dis} = -\zeta eU_a, \tag{32}$$

where

$$\zeta = \left(1 - \frac{N_1}{N}\right)\frac{E_r}{E_a}$$

is the coefficient taking into account (similar to the case of an MD, see (29)) that the energy $W_{sc} + W_{ks}$ of the domain is not dissipated.

Thus, the energy losses in the GD, as well as in the VD and MD, is taken into account by the energy $w_{kr}^{\downarrow} \propto w_a = w_{dis}$ of particles traveling relative to a collective structure (domain). Since $w_a \propto w_{kr}^{\downarrow} \propto u_d^2$, then the principle of minimum energy dissipation and the principle of minimization of Lagrangian Λ as applied to a GD give essentially the same result. However, the principle of Λ minimization is more informative, since formula (22) explains the mechanism of reduction of the drift velocity.

In a weak field, a GD is a resistor in which $\theta \cong 1$, $\zeta = 1$, and the entire energy $-eU_a$ acquired from the power supply by each particle during drift converts into heat. The weak-field mode is explained essentially similarly by the principle of minimum energy dissipation and the principle of Lagrangian minimization.

Thus, the analysis of various systems showed that, near thermodynamic equilibrium, the principle of Lagrangian minimization is practically equivalent to the principle of minimum energy dissipation. As soon as the degree of nonequilibrium increases, the principle of Λ minimization becomes more informative because the integral Lagrangian takes into account all forms of the system energy, rather than only one.

In the general form, the law of conservation of energy in the system takes the form

$$w_N = w_\Sigma + w_{kr}, \tag{33}$$

where $w_\Sigma = w_{sc} + \sum_i w_{ksi}$ is the energy of the collective form.

Let us write formula (33) in a different way, in particular, as $w_{kr} = w_N - w_\Sigma$. Now, the right-hand side contains those forms of the energy which are not dissipated, but conserved due to the energy w_{kr} of particles that take part in the irreversible process of the matter exchange with the environment. The requirement of decrease of the energy w_{kr}, which follows from the principle of Lagrangian Λ minimization, means that a structure will be formed whose existence requires the energy w_{kr}.

CONCLUDING REMARKS

Analysis of particular systems showed that the thermodynamic principles of minimum energy dissipation and minimum entropy production qualitatively correctly account for the role of dissipation in the formation of steady states of the systems in which the principle of local equilibrium is not obeyed. This is explained by the fact that the energy w_Σ of the collective structure is not dissipated. Only the energy w_{kr} of the particles traveling relative to the structure, which exists in all open systems and is basic in the systems whose states occur on the thermodynamic branch, converts into heat. In the general case, the requirement of minimization of the energy w_{kr} imposes more rigid constraints on the self-organizing system than does the requirement of the minimum of w_{dis}.

As the number of particles in the system increases, the potential energy W_{sc} of the structure grows most rapidly. If there are no restrictions on the number of particles coming from outside, the evolution of the system can be described by inequalities (16): $dw_{sc}/dt \geq 0$, $dW_{sc}/dt \geq 0$. The trend to increase in W_{sc} may be called the principle of natural evolution of open systems. The action of this principle may explain, for example, the concentration of matter in stars and galaxies, as well as the large variety of animate beings, which went the way in their development from prokaryotes to multicellular organisms of much greater size.

The principle of Lagrangian minimization claims that, among all potential systems which can be formed in an open system fed from the outside by energy and matter flows, nature will realize the structure characterized by a minimum value of the integral Lagrangian. The rejection of other possible structures and selection of only one structure satisfying the principle indicated, results in symmetry violation in nature and leads to the notions of "optimum" structure formation and its purposeful evolution.

REFERENCES

1. H. Haken, *Synergetics: An Introduction* (Springer-Verlag, Berlin, 1977).
2. M. E. Levinshteĭn, Yu. K. Pozhela, and M. S. Shur, *Gunn Effect* (Sov. Radio, Moscow, 1975) [in Russian].
3. N. A. Kervalishvili, Fiz. Plazmy **15**, 174 (1989) [Sov. J. Plasma Phys. **15**, 98 (1989)].
4. A. V. Smirnov and V. G. Usychenko, Radiotekh. Élektron. (Moscow) **36**, 151 (1991).
5. A. V. Agafonov, V. M. Fedorov, and V. P. Tarakanov, Preprint No. 37, FIAN (Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 1997).
6. V. G. Usychenko, Radiotekh. Élektron. (Moscow) **46**, 1489 (2001).
7. N. A. Kervalishvili, Fiz. Plazmy **15**, 362 (1989) [Sov. J. Plasma Phys. **15**, 211 (1989)].
8. S. V. Korobtsev, D. D. Medvedev, and V. D. Rusanov, Fiz. Plazmy **19**, 567 (1993) [Plasma Phys. Rep. **19**, 291 (1993)].
9. V. A. Vasil'ev, Yu. M. Romanovskiĭ, and V. G. Yakhno, *Autowave Processes* (Nauka, Moscow, 1987) [in Russian].
10. I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes*, 3rd ed. (Interscience, New York, 1968).
11. P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations* (Wiley, New York, 1971).
12. H. Haken, *Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices* (Springer-Verlag, New York, 1983).
13. W. Ebeling, *Strukturbildung bei Irreversiblen Prozessen* (Teubner, Leipzig, 1976).
14. D. S. Chernavskiĭ, *Synergetics and Information* (Nauka, Moscow, 2001) [in Russian].
15. W. Ebeling, A. Engel, and R. Feistel, *Physik der Evolutionsprozesse* (Akademie-Verlag, Berlin, 1990; Editorial URSS, Moscow, 2001).

16. L. A. Vaĭnshteĭn and V. A. Solntsev, *Lectures on Microwave Electronics* (Sov. Radio, Moscow, 1973) [in Russian].

17. C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, 1962).

18. Yu. P. Raizer, *Gas Discharge Physics* (Nauka, Moscow, 1992; Springer-Verlag, Berlin, 1991).

19. L. Brillouin, Phys. Rev. **60**, 385 (1941).

20. V. G. Usychenko, Radiotekh. Élektron. (Moscow) **41**, 1243 (1996).

21. O. Bahneman, in *Crossed-Field Microwave Devices*, Ed. by E. Okress (Academic, New York, 1961).

22. B. L. Usherovich, Obz. Elektron. Tekh., No. 7 (76) (1969).

23. V. G. Usychenko, Radiotekh. Élektron. (Moscow) **44**, 623 (1999).

24. R. Jepsen, in *Crossed-Field Microwave Devices*, Ed. by E. Okress (Academic, New York, 1961), Vol. 1.

25. V. G. Usychenko, Radiotekh. Élektron. (Moscow) **44**, 746 (1999).

26. M. Shur, *GaAs Devices and Circuits* (Plenum, New York, 1987).

27. N. Braslau and P. Hauge, IEEE Trans. Electron Devices **17**, 616 (1970).

28. I. Gyarmati, *Non-Equilibrium Thermodynamics* (Springer-Verlag, Berlin, 1970).

29. N. N. Moiseev, *Algorithms of Development* (Nauka, Moscow, 1987) [in Russian].

30. S. D. Gvozdover, *Theory of Electronic Microwave Devices* (GITTL, Moscow, 1956) [in Russian].

31. *Noise in Electron Devices,* Ed. by L. D. Smullin and G. A. Hause (MIT, Cambridge, 1959).

Translated by M. Lebedev

=========== **SOLIDS** ===========

Effect of Surface Properties on a Temperature Jump in a Metal

A. V. Latyshev and A. A. Yushkanov

Moscow State Regional University, ul. Radio 10a, Moscow, 105005 Russia
e-mail: latyshev@orc.ru, yushkanov@mtu-net.ru
Received December 8, 2003

Abstract—The problem of temperature jump that is induced by a heat flux toward a metal surface is solved with regard to electron energy accommodation on the surface. The temperature profile is constructed for different ratios of the electron mean free path to the Debye length. © *2004 MAIK "Nauka/Interperiodica".*

The temperature distribution near a metal surface is a key issue in the problems of metal–environment heat exchange at low temperatures. This issue is also of great importance for small-size metallic objects (the case typical of microelectronics), where processes taking place at the surface have a fundamental effect on the temperature distribution.

Consider a planar metal–environment interface and assume that a heat flux in the metal is directed toward (or opposite to) the surface. Let the Cartesian coordinate system be centered on the surface and its x axis be directed into the metal normally to the surface. Then, at distances from the surface that are much greater than electron mean free path l, the heat flux can be represented by constant temperature gradient $G_T = dT/dx$ (the metal is assumed to be isotropic). The temperature gradient is assumed to be flat in the sense that the relative temperature drop over length l is much smaller than unity.

The layer of thickness λ that is adjacent to the interface will be referred to as the Knudsen layer, as is customary in kinetic considerations. Outside the Knudsen layer, the temperature profile has the form $T = T_0 + G_T x$ for $x > 0$ and $T = T_0 + G_T x$ for $x < 0$. The value of $\Delta T = T_0 - T_s$ (where T_s is the surface temperature) will be called the temperature jump. If the gradient is flat, the temperature jump is proportional to it:

$$\Delta T = C_T l G_T. \qquad (1)$$

Coefficient C_T, which is independent of G_T, is called the temperature jump coefficient. It is found by solving the kinetic equation in the Knudsen layer [1]. Usually, researchers deal with dimensionless temperature jump $\varepsilon_T = C_T l g_T$, where $g_T = G_T/T_s$.

Note that the temperature jump significantly depends on the type of interaction between electrons and metal the surface. Taking account of this fact requires that the boundary conditions be modified so that they include the electron energy accommodation at

the interface. To this end, an appropriate accommodation coefficient should be introduced.

We consider the general case of the arbitrary degeneration of the electron gas. Therefore, the results obtained are valid in a wide temperature range.

Consider a metal for which the Fermi surface is spherical. For electrons in metals, a kinetic equation in the τ approximation [1–3] is frequently used:

$$\frac{\partial f}{\partial t} = (\mathbf{v} \cdot \nabla)f + e_0 \mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{p}} = \frac{1}{\tau}(f_F^0 - f). \qquad (2)$$

Here, f is the electron distribution function, e_0 is the electron charge, $\mathbf{p}$ is the electron momentum, $\mathbf{E}$ is the electric field, $\mathbf{v}$ is the electron velocity, τ is the electron relaxation time, and f_F^0 is the Fermi distribution function. If electron scattering by impurities prevails, it is reasonable to assume that the electron free path is constant [4]. In this case, the relaxation time can be expressed as $\tau = l/w$. Here, l is the electron free path and $w = |\mathbf{v} - \mathbf{u}|$, where $\mathbf{u}$ is the mean electron velocity. Note that, in metals, the inequality $u \ll v$ is valid in all physically feasible conditions. Therefore, we hereafter will use v instead of w.

Note also that, generally, the kinetic equation for phonons, along with that for electrons, should be taken into consideration when the temperature is finite.

Let phonon distribution function f_{ph} satisfy the kinetic equation [1, 4]

$$\frac{\partial f_{ph}}{\partial t} + (\mathbf{C} \cdot \nabla)f_{ph} = J(f_{ph}, f), \qquad (3)$$

where $\mathbf{C}$ is the phonon velocity and $J(f_{ph}, f)$ is the collision integral including electron–phonon, electron–impurity, and phonon–phonon scattering. For a finite temperature, the Fermi distribution function with some effective temperature T_* and effective chemical potential μ_* must appear in Eq. (2) as the equilibrium elec-

tron distribution function instead of f_F^0, which corresponds to the zero temperature. It is noteworthy that these effective temperature and chemical potential generally differ from the locally equilibrium values of the temperature, T_{eq}, and chemical potential, μ_{eq}.

Then, the kinetic equation for electrons takes the form

$$\frac{\partial f}{\partial t} + (\mathbf{v} \cdot \nabla) f + e_0 \mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{p}} = \frac{v}{l}(f_F^* - f). \quad (4)$$

Here, $f_F^* = f_F(\mu_*, T_*) = f_F(\mu_*, T_*) = \left[\exp\left(\frac{mv^2}{2kT_*} - \frac{\mu_*}{kT_*}\right) + 1\right]^{-1}$ is the Fermi distribution function (Fermian), k is the Boltzmann constant, and m is the electron effective mass.

The electron conservation law leads to the following equation as a consequence of Eq. (4):

$$\int vf \, d\Omega_F = \int v f_F^* \, d\Omega_F, \quad (5)$$

where $d\Omega_F = (2s + 1)(2\pi\hbar)^{-3} d\mathbf{p}$, $\hbar$ is the Planck constant, and s is the electron spin. Integration is carried out over the entire velocity space. In most metals, the electron subsystem makes a major contribution to heat transfer [4]. We will consider just this case and ignore the phonon contribution to this process.

In the steady-state case, kinetic equation (4) for electrons is recast as

$$\nabla \cdot \mathbf{Q}_e = \frac{v}{l}\left(\int v \frac{m}{2} v^2 f_F^* d\Omega_F - \int v \frac{m}{2} v^2 f \, d\Omega_F\right), \quad (6)$$

where $\mathbf{Q}_e$ is the heat flux due to electrons.

In the steady-state case, where heat sources are absent, total heat flux $\mathbf{Q}$ is constant; that is, $\nabla \cdot \mathbf{Q} = 0$. If the phonon contribution to the heat transfer process in a metal can be neglected, we also assume, in accordance with (6), that $\nabla \cdot \mathbf{Q}_e = 0$. Then, from Eq. (5), we get

$$\int v \frac{m}{2} v^2 f \, d\Omega_F = \int v \frac{m}{2} v^2 f_F^* d\Omega_F. \quad (7)$$

Relationships (5) and (7) specify parameters T_* and μ_* appearing in kinetic equation (4).

In the approximation discussed thus far, Eq. (4) for electrons and Eq. (3) for phonons turn out to be independent. In this case, a number of fine effects are missed. To take them into account, it is necessary to employ the more adequate τ approximation that is considered in this work.

Let us assume that characteristic temperature drops over length l are small compared with the electron gas temperature. Then, the problem can be linearized. The

distribution function will be sought in the form $f = f_F^s + \varphi(t, \mathbf{r}, \mathbf{v})g$. Here, $f_F^s = f_F(\mu_s, T_s)$, φ is a new unknown function, μ_s is the chemical potential of the electrons scattered by the surface, $g = \partial f_F^s / \partial \varepsilon_s$, $\varepsilon_s = (mv^2/2 - \mu_s)/kT_s$, and T_s is the surface temperature.

Let us introduce the designations

$$\mathbf{c} = \sqrt{\frac{m}{2kT_s}} \mathbf{v}, \quad \alpha = \frac{\mu_s}{kT_s}, \quad \varepsilon_* = \frac{mv^2}{2kT_*} - \frac{\mu_*}{kT_*}.$$

In terms of these designations,

$$f_F^*(\varepsilon_*) = \frac{1}{\exp(\varepsilon_*) + 1}, \quad f_F^s(c, \alpha) = \frac{1}{\exp(c^2 - \alpha) + 1},$$

$$g = g(c, \alpha) = \frac{\exp(c^2 - \alpha)}{[\exp(c^2 - \alpha) + 1]^2}.$$

Now, we pass to the dimensionless variables to linearize local Fermian f_F^*. Note that

$$\varepsilon_* = \frac{T_s}{T_*}\left[\frac{m}{2kT_s}v^2 - \frac{\mu_*}{kT_s}\right] = \frac{T_s}{T_*}(c^2 - \alpha_*), \quad \alpha_* = \frac{\mu_*}{kT_s}.$$

Since $T_* = T_s + \delta T_s$ and $\alpha_* = \alpha + \delta\alpha_*$, we come to

$$\varepsilon_* = c^2 - \alpha - \frac{\delta T_*}{T_s}(c^2 - \alpha) - \delta\alpha_*,$$

hence,

$$\delta\varepsilon_* = -\delta\alpha_* - (c^2 - \alpha)\frac{\delta T_*}{T_s},$$

where $\delta\varepsilon_* = \varepsilon_* - \varepsilon_s$ and $\varepsilon_s = c^2 - \alpha$.

Consequently,

$$f_F^* = f_F^s + \left(\frac{\partial f_F^*}{\partial \varepsilon_*}\right)_{\varepsilon_* = \varepsilon_s} \delta\varepsilon_*$$

or

$$f_F^* = f_F^s + g(c, \alpha)\left[\delta\alpha_* + (c^2 - \alpha)\frac{\delta T_*}{T_s}\right].$$

Let us introduce dimensionless quantities $\mathbf{e} = (e_0 l / kT_s)\mathbf{E}$, $t_* = t\sqrt{2kT_s/m}/l$, and $\mathbf{r}_* = \mathbf{r}/l$ (hereafter, the asterisks will be omitted). In terms of the dimensional variables, Eq. (4) takes the form

$$\frac{\partial\varphi}{\partial t} + (\mathbf{c} \cdot \nabla)\varphi - \mathbf{c} \cdot \mathbf{e} = c\left[\delta\alpha_* + (c^2 - \alpha)\frac{\delta T_*}{T_s} - \varphi\right]. \quad (8)$$

The parameters of this equation, $\delta\alpha_*$ and δT_*, are found from the conservation laws (relationships (5) and

(7)), which are now written as

$$\int \left[\delta\alpha_* + (c^2 - \alpha)\frac{\delta T_*}{T_s} - \varphi \right] c^n g d^3 c = 0; \quad n = 1, 3.$$

From this system, we find

$$\frac{\delta T_*}{T_s} = -\frac{r_1(\alpha)}{2\pi\Delta(\alpha)}\int \varphi c g d^3 c + \frac{l(\alpha)}{2\pi\Delta(\alpha)}\int \varphi c^3 g d^3 c,$$

$$\delta\alpha_* - \alpha\frac{\delta T_*}{T_s} = \frac{r_3(\alpha)}{2\pi\Delta(\alpha)}\int \varphi c g d^3 c \quad (9)$$

$$- \frac{r_1(\alpha)}{2\pi\Delta(\alpha)}\int \varphi c^3 g d^3 c.$$

Here,

$$r_1(\alpha) = 4\int_0^\infty c\ln[1 + \exp(\alpha - c^2)]dc,$$

$$l(\alpha) = \ln(1 + \exp(\alpha)),$$

$$r_3(\alpha) = 12\int_0^\infty c^3\ln[1 + \exp(\alpha - c^2)]dc,$$

$$\Delta(\alpha) = l(\alpha)r_3(\alpha) - r_1^2(\alpha).$$

Let us represent Eq. (8) in the conventional form:

$$\frac{\partial\varphi}{\partial t} + c\varphi(t, \mathbf{r}, \mathbf{c}) - \mathbf{ce}(\mathbf{r})$$

$$= \frac{c}{2\pi}\int k(c, c')\varphi(t, \mathbf{r}, \mathbf{c}')d\Omega(\alpha),$$ (10)

where

$$k(c, c') = 1 + \frac{l^2(\alpha)}{\Delta(\alpha)}\left(c^2 - \frac{r_1(\alpha)}{l(\alpha)}\right)\left(c'^2 - \frac{r_1(\alpha)}{l(\alpha)}\right),$$

$$d\Omega(\alpha) = \frac{g(c', \alpha)c'}{l(\alpha)}d^3 c'.$$

Let the half-space $x > 0$ be filled with a metal and plane xy be coincident with the metal boundary. We assume that a heat flux present in the metal is directed normally to the surface. Then, the heat-flux-induced electric field will also be directed normally to the surface, so that all the parameters of the problem will depend only on coordinate x.

The set of equations that describe the problem consists of Eq. (10) for electrons and the equation for the electric field. In dimensionless form, these equations look like

$$\mu\frac{\partial\varphi}{\partial x} + \varphi(x, \mu, c) - \mu e(x)$$

$$= \frac{1}{l(\alpha)}\int_{-1}^{1}\int_0^\infty k(c, c')\varphi(x, \mu', c')g(c')c'^3 d\mu' dc',$$

(11)

$$e'(x) = a_0^2\int_{-1}^{1}\int_0^\infty \varphi(x, \mu, c)g(c)c^2 d\mu dc,$$

$$\mu = \frac{c_x}{c}, \quad a_0^2 = \frac{e_0^2 m^2 l^2}{\pi^2\varepsilon_0\hbar^3}\sqrt{\frac{2kT_s}{m}},$$

where ε_0 is the dielectric constant.

It should be emphasized that to state boundary conditions for electrons on a metal surface is a challenge [4]. The specular–diffuse Maxwell equation is frequently used. However, it is impossible to relate all scattering characteristics of electrons to a single parameter that shows to which extent the scattering is specular. Even if the scattering is perfectly diffuse, there is the possibility that the electron energy will remain unchanged. For electron scattering on the surface, heat transfer can be described by means of additional coefficient q_e that takes into account energy accommodation. Consider perfectly diffuse electron scattering on the surface. In this case, the distribution function of the electrons reflected by the surface coincides with the Fermi equilibrium distribution $f_F^e = \exp[(\varepsilon_e) + 1]^{-1}$, where $\varepsilon_e = (mv^2/2 - \mu_e)/kT_e$. In general, parameters T_e and μ_e differ from surface parameters T_s and μ_s. They coincide only if accommodation coefficient q_e equal unity.

In the linear approximation, distribution function f_F^e can be represented as

$$f_F^e = f_F^s + g(c, \alpha)\left[\delta\alpha_e + (c^2 - \alpha)\frac{\delta T_e}{T_s}\right],$$

where $\delta\alpha_e = (\mu_e - \mu_s)/kT_s$ and $\delta T_e = T_e - T_s$.

We find that the boundary conditions including the accommodation character of electron scattering on the surface [3] and the conditions inside the metal have the form

$$\varphi(0, \mu, c) \equiv \varphi_0(\mu, c) = A_0 + (c^2 - r_3(\alpha)/r_1(\alpha))B_0,$$

$$0 < \mu < 1,$$

$$\varphi(x, \mu, c) = \varphi_{as}(x, \mu, c) + o(1), \quad (12)$$

$$x \longrightarrow +\infty, \quad -1 < \mu < 0,$$

$$e(0) = 0, \quad e(x) = e_{as} + o(1), \quad x \longrightarrow +\infty.$$

Here,

$$\varphi_{as} = \varepsilon_T\left(c^2 - \frac{r_0(\alpha)}{s(\alpha)}\right)$$

$$+ g_T(x - \mu)\left(c^2 - \frac{r_0(\alpha)}{s(\alpha)}\right) - h_0(\alpha)g_T\mu,$$

$$e_{as} = -g_T h_0(\alpha), \quad \Delta_0(\alpha) = r_0(\alpha)l(\alpha) - s(\alpha)r_1(\alpha),$$

$$h_0(\alpha) = \frac{\Delta_0(\alpha)}{s(\alpha)l(\alpha)} = \frac{r_0(\alpha)}{s(\alpha)} - \frac{r_1(\alpha)}{l(\alpha)},$$

$$B_0 = \frac{T_e - T_s}{T_s},$$

$$r_0(\alpha) = \frac{2}{3}\int_0^\infty \ln[1 + \exp(\alpha - c^2)]dc,$$

$$s(\alpha) = \int_0^\infty \frac{\exp(\alpha - c^2)}{1 + \exp(\alpha - c^2)}dc.$$

It is assumed that the energy flux E_r of the electrons reflected from the surface is related to the energy flux E_i of the electrons incident on the surface as

$$q_e = (E_i - E_r)(E_i - E_s)^{-1}. \tag{13}$$

Here,

$$E_i = -\int_{-1}^0\int_0^\infty \varphi(0, \mu, c)c^5\mu g(c)d\mu dc,$$

$$E_r = \int_0^1\int_0^\infty \varphi(\mu, c)c^5\mu g(c)d\mu dc,$$

E_s is the electron energy flux from the wall that corresponds to the thermodynamic equilibrium between the wall and electrons ($T_e = T_s$), and

$$E_s = \int_0^1\int_0^\infty \varphi_s(\mu, c)c^5\mu g(c)d\mu dc,$$

with the proviso that $\varphi_s(\mu, c) = A_s$ for $0 < \mu < 1$. Quantity A_s is determined from the condition

$$\int_0^1\int_0^\infty \varphi_s(\mu, c)c^3\mu g(c)d\mu dc$$

$$= -\int_{-1}^0\int_0^\infty \varphi(0, \mu, c)c^3\mu g(c)d\mu dc.$$

This condition can be written in terms of the electron flows; that is, $N_s = N_i$, where N_s is the electron flux from the wall that corresponds to the thermodynamic

equilibrium between the wall and electrons and N_i is the electron flow toward the wall.

Parameter q_e, varying between zero and unity, is called the energy accommodation coefficient. It is an empirical quantity. Its value depends on the type of interaction between electrons, surface phonons, and the material adjacent to the metal. At low temperatures and in the case where the metal is in contact with an allowable substance (gas), q_e may be considerably smaller than unity, since heat exchange between the electrons and other components is difficult under these conditions. Thus, taking into account energy accommodation in describing heat exchange processes at the interface is of great significance.

Along with the energy accommodation condition on the surface, one more condition, namely, the no-percolation condition, must be met. Mathematically, this means that total electron flow N_0 across the interface is zero:

$$N_0 = \int_{-1}^1\int_0^\infty \varphi(0, \mu, c)c^3\mu g(c)d\mu dc = 0,$$

or, which is the same, $N_0 = N_r - N_i = 0$, where

$$N_r = \int_0^1\int_0^\infty \varphi(0, \mu, c)c^3\mu g(c)d\mu dc$$

is the flow of the electrons reflected from the wall.

Similarly, total electron energy flux E_0 can be represented as $E_0 = E_r - E_i$. Then, the electron energy accommodation condition can be recast as $q_e(E_i - E_s) + E_0 = 0$, where the total energy flux

$$E_0 = \int_{-1}^1\int_0^\infty \varphi(x, \mu, c)c^5\mu g(c)d\mu dc$$

is the same for all $x > 0$ by virtue of the energy conservation law. The value of E_0 is determined through the asymptotic distribution function at the wall:

$$E_0 = \int_{-1}^1\int_0^\infty \varphi_{as}(0, \mu, c)c^5\mu g(c)d\mu dc = -\frac{1}{3}\frac{\Delta(\alpha)}{l(\alpha)}g_T.$$

It is easy to check that $E_s = r_1(\alpha)A_s/4$ and $E_r = r_1(\alpha)A_0/4$; therefore,

$$E_i = E_r - E_0 = \frac{1}{4}r_1(\alpha)A_0 + \frac{1}{3}\frac{\Delta(\alpha)}{l(\alpha)}g_T,$$

$$E_i - E_s = E_r - E_0 - E_s$$

$$= \frac{1}{4}r_1(\alpha)(A_0 - A_s) + \frac{1}{3}\frac{\Delta(\alpha)}{l(\alpha)}g_T.$$

Then, from the electron energy accommodation condition, we find that

$$A_0 = A_s + \frac{4}{3}\frac{1-q_e}{q_e}\frac{\Delta(\alpha)}{l(\alpha)r_1(\alpha)}g_T. \qquad (14)$$

From the relationships for the electron flows, $N_s = N_i$ and $N_r = N_i$, it follows that $N_s = N_r$; hence,

$$A_s = A_0 - \frac{\Delta(\alpha)}{l(\alpha)r_r(\alpha)}B_0. \qquad (15)$$

From (14) and (15), we find quantity B_0:

$$B_0 = \frac{4}{3}\frac{1-q_e}{q_e}g_T.$$

Thus, we have found constant B_0 for the first boundary condition in (12). The second constant, A_0, is not specified beforehand: it is determined from a solution to the problem.

Function φ_{as} describes the thermal conductivity of the electron gas inside the metal. The heat conduction process generates electric field e_{as} (the thermoelectric effect described, e.g., in [1]). The condition $e(0) = 0$ stems from the assumption that an electric field outside the metal is absent.

It should be stressed that functions φ_{as} and e_{as} are solutions to set (11).

According to the structure of function φ_{as}, we seek function φ in the form

$$\varphi(x, \mu, c) = h_1(x, \mu) + \left(c^2 - \frac{r_0(\alpha)}{s(\alpha)}\right)h_2(x, \mu) \qquad (16)$$

and arrive at the set of equations

$$\mu\frac{\partial h_1}{\partial x} + h_1(x, \mu) = \bar{h}_1(x) + \mu e(x), \qquad (17)$$

$$\mu\frac{\partial h_2}{\partial x} + h_2(x, \mu) = \bar{h}_2(x),$$

$$\bar{h}_j(x) = \frac{1}{2}\int_{-1}^{1}h_j(x, \mu)d\mu; \quad j = 1, 2, \qquad (18)$$

$$e'(x) = a^2\bar{h}_1(x), \quad a^2 = a_0^2 s(\alpha). \qquad (19)$$

In view of (16), boundary conditions (12) have the form

$$h_1(0, \mu) = A_0 + l_1(\alpha)B_0, \quad 0 < \mu < 1,$$

$$l_1(\alpha) = \frac{\Delta_1(\alpha)}{s(\alpha)r_1(\alpha)} = \frac{r_0(\alpha)}{s(\alpha)} - \frac{r_3(\alpha)}{r_1(\alpha)},$$

$$\Delta_1(\alpha) = r_0(\alpha)r_1(\alpha) - s(\alpha)r_3(\alpha), \qquad (20)$$

$$h_1(x, \mu) = -h_0(\alpha)g_T\mu + o(1),$$

$$x \longrightarrow \infty, \quad -1 < \mu < 0,$$

$$h_2(0, \mu) = B_0, \quad 0 < \mu < 1,$$

$$h_2(x, \mu) = \varepsilon_T + g_T(x - \mu) + o(1), \quad x \longrightarrow \infty, \qquad (21)$$

$$-1 < \mu < 0,$$

$$e(0) = 0, \quad e(x) = e_{as} + o(1), \quad x \longrightarrow +\infty. \qquad (22)$$

We draw the reader's attention to the fact that the general problem stated by (11) and (12) is split into two subproblems that are interrelated only through the boundary conditions. The first subproblem, (18) and (21), is associated with a temperature jump in a metal; the second one stated by (17), (19), (20), and (21) is related to the behavior of an electric field near the surface in the presence of temperature gradient g_T normal to the surface.

Consider first the subproblem stated by (18) and (21). According to [5], a solution to this subproblem is looked for in the form of the expansion in eigensolutions:

$$h_2(x, \mu) = \varepsilon_T + g_T(x - \mu)$$

$$+ \int_0^1 \exp\left(-\frac{x}{\eta}\right)\Phi(\eta, \mu)m(\eta)d\eta. \qquad (23)$$

In expansion (23), the unknowns are temperature jump ε_T and function $m(\eta)$.

Omitting the solution of the problem (the solution method is given elsewhere [5, 6]), we give the expressions for these unknowns:

$$\varepsilon_T = V_1 g_T + B_0 = \left(V_1 + \frac{3(1-q_e)}{4q_e}\right)g_T,$$

$$V_1 = \frac{1}{\pi}\int_0^1\zeta(u)du = 0.71045\ldots, \qquad (24)$$

$$\eta m(\eta) = -\frac{g_T}{\pi i}\left[\frac{1}{X^+(\eta)} - \frac{1}{X^-(\eta)}\right] = \frac{2g_T\sin\zeta(\eta)}{\pi X(\eta)},$$

where

$$X(z) = \frac{1}{z}\exp V(z), \quad V(z) = \frac{1}{\pi}\int_0^1\frac{\zeta(u)du}{u - z},$$

$$\zeta(u) = -\frac{\pi}{2} - \arctan\frac{2\lambda(u)}{\pi u}.$$

Formula (24) specifies the desired temperature jump in the metal. Comparing formulas (1) and (24), we determine the temperature jump coefficient:

$$C_T = V_1 + \frac{3(1-q_e)}{4q_e}.$$

This coefficient does not depend on the electron gas degeneracy. It follows from this formula that the temperature jump grows with decreasing accommodation

$\delta T_*(x)/T_g g_T$
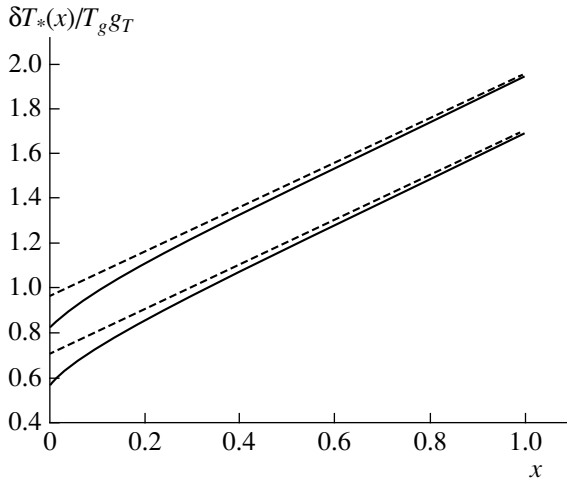
Temperature profiles.

coefficient. For $q_e = 1$, this formula yields the well-known expression for temperature jump that was derived in [6].

The temperature profile in the metal will be constructed in accordance with (9). From expansion (16), we get

$$\frac{1}{2\pi}\int \varphi g c d^3 c = l(\alpha)\bar{h}_1 - \frac{\Delta_0(\alpha)}{s(\alpha)}\bar{h}_2,$$

$$\frac{1}{2\pi}\int \varphi g c^3 d^3 c = r_1(\alpha)\bar{h}_1 - \frac{\Delta_1(\alpha)}{s(\alpha)}\bar{h}_2.$$

Substituting these equalities into (9) yields

$$\frac{\delta T_*(x)}{T_s} = \bar{h}_2 \equiv \frac{1}{2}\int_0^1 h_2(x,\mu)d\mu$$

$$= \varepsilon_T + g_T x + \frac{1}{2}\int_0^1 \exp\left(-\frac{x}{\eta}\right)m(\eta)d\eta,$$

or

$$\frac{\delta T_*(x)}{T_s} = \varepsilon_T + g_T x + g_T \frac{1}{\pi}\int_0^1 \exp\left(-\frac{x}{\eta}\right)\frac{\sin\zeta(\eta)}{X(\eta)}\frac{d\eta}{\eta}.$$

On the wall ($x = 0$), this expression simplifies to

$$\frac{\delta T_*(x)}{T_s} = \varepsilon_T + g_T\left(-V_1 + \frac{1}{X(+0)}\right)$$

$$= g_T\left(\frac{1}{\sqrt{3}} + \frac{4(1-q_e)}{q_e}\right).$$

The figure demonstrates the temperature profiles in the half-space $x > 0$ that are constructed based on the analytical solution found. The dashed lines show the asymptotic profiles extrapolated to the boundary of the metal; the solid lines, the actual temperature profiles. The upper curves correspond to the accommodation coefficient $q_e = 0.75$; the lower ones, to $q_e = 1$. It is seen that the temperature profile shifts upward as the accommodation coefficient decreases.

ACKNOWLEDGMENTS

REFERENCES

1. L. D. Landau and E. M. Lifshits, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
2. A. A. Abrikosov, *Introduction to the Theory of Metals* (Nauka, Moscow, 1987) [in Russian].
3. J. M. Ziman, *Electrons and Phonons* (Clarendon, Oxford, 1960; Inostrannaya Literatura, Moscow, 1962).
4. M. M. R. Williams, *Mathematical Methods in Particle Transport Theory* (Butterworth, London, 1971).
5. A. V. Latyshev and A. A. Yushkanov, Fiz. Tverd. Tela (St. Petersburg) **43**, 1744 (2001) [Phys. Solid State **43**, 1816 (2001)].
6. A. V. Latyshev and A. A. Yushkanov, Zh. Tekh. Fiz. **73** (7), 37 (2003) [Tech. Phys. **48**, 837 (2003)].

Translated by V. Isaakyan

=== **SOLIDS** ===

Criterion of the Incubation Time in the Problems of Pulsed Fracture and Electric Breakdown

Yu. V. Petrov and P. A. Glebovskiĭ

St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 198904 Russia

e-mail: yp@yp1004.spb.edu, pietr@mail.ru

Received February 3, 2004

Abstract—The problems of a pulsed strength of continuum media are considered in terms of the structural–time approach that is based on the concept of the incubation fracture time. This approach makes it possible to describe phenomena that arise under high-velocity external effects. A limiting condition that determines the instant of rupture or breakdown is proposed on the basis of the structural–time approach. A way to interpret and to determine the incubation time is proposed. A phenomenological model of an electric breakdown of solid dielectrics is formulated. Examples are considered where the structural–time approach is applied to problems of spall fracture, crack initiation, and a pulsed breakdown of dielectrics. A procedure for describing the time dependence of the electric strength (volt–second characteristic) is described in detail. The results of the calculations are found to be in good agreement with experimental data. © 2004 MAIK "Nauka/Interperiodica".

Investigation of dynamic fracture of solids and insulators under the high-rate (pulsed) effect of the ambient medium under various conditions is of great interest from the point of view of continuum physics and mechanics. The results of such investigations are used in creating and operating various devices in machine building and power engineering.

Experiments to study dynamic fracture reveal a number of effects that show a drastic distinction between a fast dynamic rupture (breakdown) of materials and a similar process under quasistatic conditions. For example, the dependence of limiting characteristics on history and on the way in which the load is applied is one of the main problems in exploring dynamic fracture. This effect manifests itself in almost all of situations of fast fracture. By way of example, we indicate that, in the case of the fracture of a defect-free continuum, one defines the limiting characteristic as the minimum amplitude of a stress pulse that leads to the rupture of the material being considered. For the intensity of a local force field, this characteristic provides a limit above which there occurs fracture. Attempts at determining, for a local field, a critical intensity that would correspond to specific rates of loading lead to ambiguous results. The dependence on the way in which an external force is applied manifests itself as a change in the limiting values in response to a change in a number of factors, including the duration of its action, its amplitude, and the rate of its growth. In the case of a quasistatic external force, the critical value is a constant peculiar to a given material, while, in the case of dynamic fracture, the values determined experimentally for critical characteristics are highly unstable, with the result that their behavior proves to be unpredictable.

The above features in the behavior of materials (as well as some other special features of their behavior) that are subjected to pulsed external forces appear to be common to apparently different physical processes such as the dynamic fracture of solids and electric breakdown in solid insulators. In the present study, we consider examples that illustrate dynamic effects in these physical processes. A universal interpretation of the fracture of solids and of electric breakdown in insulators is proposed within a structural–time approach [1, 2] that is based on the concept of incubation time.

The absence of an adequate limiting condition that would pinpoint the instant of fracture or breakdown is the main reason behind the difficulties encountered in simulating the aforementioned effects of mechanical and electric strength. This problem can be solved on the basis of macroscopic structural fracture mechanics and the concept of the incubation fracture time, which takes into account kinetic processes leading to the formation of macroscopic discontinuities [1, 2]. The above dynamic effects become significant when one goes over from slowly varying external forces to those whose periods are commensurate with the scale that is specified by the incubation fracture time. The nature of the incubation time is associated with preparatory relaxation processes involving the development of microscopic defects in the structure of a material.

The incubation-fracture-time criterion proposed in [1, 2] makes it possible to calculate effects that are associated with unstable behavior of dynamic strength characteristics and are observed in experiments studying the fracture of solids. In a generalized form, this cri-

terion can be represented as

$$\frac{1}{\tau} \int_{t-\tau}^{t} \left(\frac{F(t')}{F_c}\right)^{\alpha} dt' \le 1, \tag{1}$$

where $F(t)$ is the intensity of the local force field that causes the fracture of a medium, F_c is the static limit of the local force field, and τ is an incubation time that is related to the dynamics of the relaxation process preparing the rupture; the fracture time is determined by the instant at which the condition in (1) reduces to a strict equality, the parameter α characterizing the sensitivity to the level of the strength of the force field that causes fracture.

For the example of the mechanical fracture of a material, we will now consider one possible way to interpret and to determine the parameter τ. We assume that a standard sample from this material is stretched and that it is ruptured into two parts at a stress P arising at some time instant taken for zero, $t = 0$; that is, $F(t) =$

Fig. 1. Time dependences of (a) the spall strength of aluminum [3, 4] (t_* is the fracture time) and (b) the dynamic fracture viscosity for crack initiation in Hormalite-100 [5] (t_* is the fracture time, and K_{1q} is the critical intensity coefficient).

$PH(t)$, where $H(t)$ is a Heaviside step function. In the case of a quasibrittle fracture, there occurs the removal of the load, with the result that, at the locus of the rupture, the local stress decreases rapidly (but not instantly) from P to zero. Concurrently, the corresponding load-removing wave is generated in the sample, and its propagation can be recorded by standard (for example, interferometric) methods. The evolution of the stress at the locus of rupture can be schematically represented as $\sigma(t) = P - Pf(t)$, where the function $f(t)$ changes from zero to unity within a time interval T. The case of $f(t) = H(t)$ corresponds to the classical theory of strength; that is, rupture is instantaneous within the classical approach ($T = 0$). But, in fact, the rupture of a material (sample) is a process that is extended in time, the function $f(t)$ describing the kinetics of a transition from a state that is thought to be defect-free [$f(0) = 0$] to a state in which the sample is fully destroyed at the point in question [$f(T) = 1$].

Applying fracture criterion (1) to this situation, we obtain $T = \tau$ at $P = F_c$; that is, the incubation time introduced above is the time period to fracture occurring as soon as the stress in the material reaches the static tensile strength. In experiments studying the static fracture of samples, this period can be measured by various methods. For example, this can be done by measuring the time of pressure growth at the front of the load-removing wave recorded with the aid of the interferometric method by using the velocity profile of points at the surface of the sample.

Further, we consider examples of how the criterion in (1) is realized in various problems of physics and mechanics.

(1) The experimental time dependence of the strength in the case of spall fracture of solids (see Fig. 1) provides an example that illustrates the intricate behavior of the strength of solids [3]. This dependence of the fracture time t_* on the threshold amplitude of a pulse P_* at various values of its duration demonstrates that the dynamic strength is not a constant characterizing a material, but that it also depends on the fracture time. The classical critical-stress criterion $\sigma(t) \le \sigma_c$, where σ_c is the dynamic strength, describes well quasistatic fracture at large times that is caused by long wave pulses $\sigma(t) = Pf(t)$, where P is the amplitude of a pulse and $f(t)$ is its time-profile function. In the case of short pulses, however, there arises a weak threshold-amplitude dependence of the fracture time, this dependence featuring an asymptote. This phenomenon is referred to as the effect of the dynamical branch in the time dependence of the strength [3].

The dynamic-branch phenomenon has not yet been explained either within classical theories of the strength or within time criteria known so far.

The total time dependence of the strength can be obtained on the basis of incubation-time criterion (1),
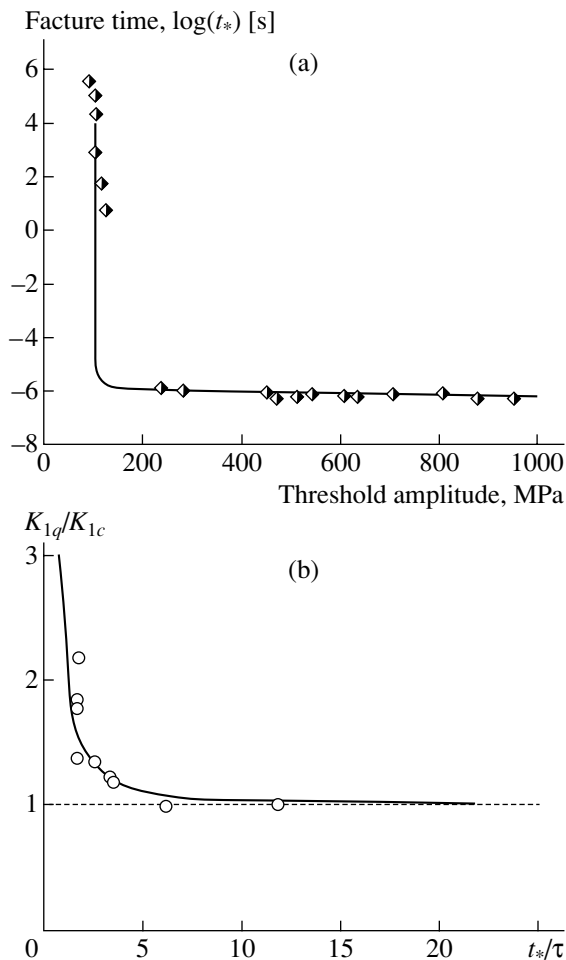
which, in our case of spall fracture, assumes the form

$$\int_{t-\tau}^{t} \sigma(t')dt' \le \sigma_c \tau, \qquad (2)$$

where $\sigma(t')$ is the time dependence of the stress at the locus of rupture.

Our calculation was performed for triangle pulses, which were used in the experiment reported in [3], and for the aluminum parameters of $\tau = 0.75$ μs and $\sigma_c = 103$ MPa. The time dependence obtained in [4] for the strength of aluminum [3] according to the criterion in (2) is represented by the solid curve in Fig. 1a.

(2) Let us consider the case where an elastic plate containing a crack is loaded symmetrically on the two sides of the crack with a uniform pressure, which increases linearly within the time t_0, whereupon it is maintained at a constant level of P; that is, the components of the stress tensor at the edges of the crack are specified in the form

$$\sigma_y = P[tH(t) - (t-t_0)H(t-t_0)]/t_0, \quad \sigma_{xy} = 0,$$

where $H(t)$ is a Heaviside step function.

The corresponding experiment was implemented in [5–7]. By changing the rate of the application of the pressure within the interval t_0 of its growth, the authors of those studies could obtain different values of the fracture time t_* and measure the starting values of the intensity coefficient (dynamic fracture viscosity) $K_{Iq} = K_I(t_*)$ that correspond to the start of growth of the crack. One result obtained in [5] for a Homalite-100 glasslike polymer is shown in Fig. 1b. The experiment exhibits the growth of the threshold intensity coefficient as the fracture time decreases; that is, as one increases the rate of the application of a load. Concurrently, the resulting velocity dependences of the dynamic fracture viscosity are highly unstable and can change sizably in response to variations in the duration of the stage within which the load grows, in the shape of the time profile of a loading pulse, in the geometry of the samples used, and in the way in which the load is applied. For example, the results obtained in [6, 7] show velocity (time) dependences of K_{Iq}, which are markedly different for the same materials, and reveal that the dynamic fracture viscosity may depend non-monotonically on the time of crack initiation under the effect of a shock wave.

Experimental results show that the dynamic fracture viscosity is not a characteristic of a material and that the introduction of the threshold-intensity-coefficient criterion $K_I(t) \le K_{Iq}$ in the theory and, accordingly, of the quantity K_{Iq} as a matter parameter that determines dynamic fracture (by analogy with the static parameter K_{Ic}) is not correct.

The behavior of the dynamic fracture viscosity can be explained and calculated on the basis of incubation-

time criterion (1), which, in the case being considered, assumes the form [2]

$$\frac{1}{\tau}\int_{t-\tau}^{t} K_I(t')dt' \le K_{Ic}. \qquad (3)$$

The results of the calculation performed in [2] for experimental data reported in [5] are represented by the solid curve in Fig. 1b. The parameters used in that calculation were set to $K_{Ic} = 0.48$ MPa $\sqrt{m}$ and $\tau = 9$ μs.

(3) A pulsed electric breakdown of solid dielectrics provides yet another example of phenomena where one can observe the effects described above. By way of example, we indicate that, in the case where the time within which the applied voltage is operative is small, the breakdown voltage usually increases as this time is decreased. In alkali halide crystals, the breakdown channel caused by an electric field applied for about 10 ns arises at a voltage several times as great as the quasistatic breakdown voltage (that is, a voltage applied within a period not shorter than 1 μs [8]). This effect was observed in the breakdown of a number of materials. As an illustration, the dependence of the breakdown electric field E^* for an ammonium perchlorate single crystal on the duration t_0 of the leading edge of a pulse is shown in Fig. 2 according to [9]. This dependence, which also characterizes the dependence of the electric strength on the rate of growth of the voltage in the sample being studied, can be referred to as the time dependence of the electric strength, being analogous to that which is observed in the above case of the dynamic fracture of materials (see Fig. 1). In the experiments described in [9, 10], thin plates from ammonium perchlorate single crystals were placed in a pulsed electric field. The electric breakdown of samples
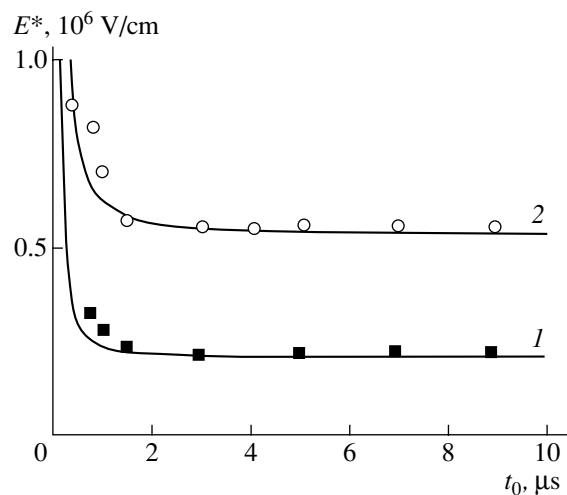

Fig. 2. Calculated dependence (solid curve) of the electric strength E^* of ammonium perchlorate on the duration of the leading edge of a pulse for various values of the distance between the electrodes: (*1*) 0.03 and (*2*) 0.01 cm.
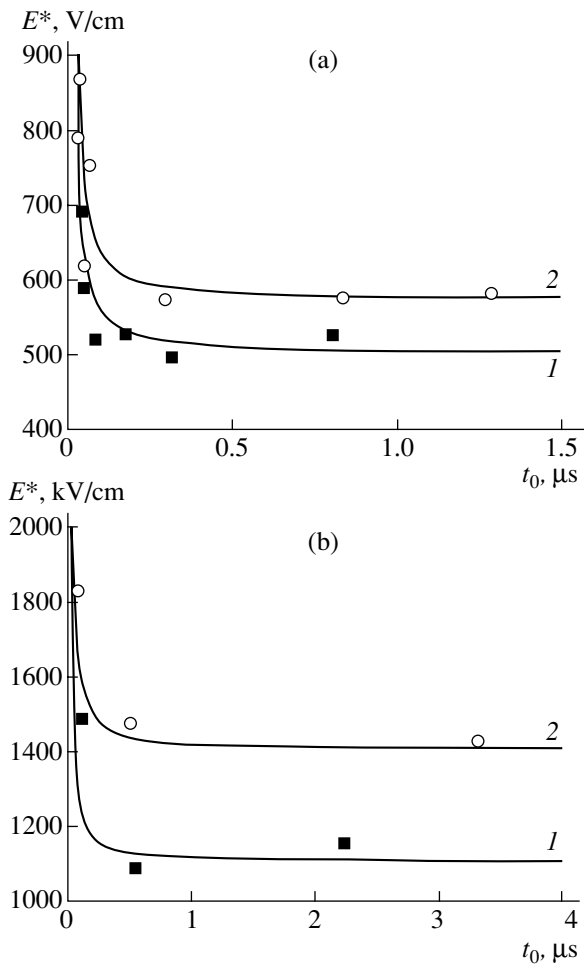
Fig. 3. (a) Calculated dependence (solid curves) of the electric strength E^* of porcelain on the duration of the leading edge of a pulse in a field close to a uniform one (distilled water was taken for a medium) for various values of the distance between the electrodes: (*1*) 1.8 and (*2*) 1 mm. The displayed experimental data were borrowed from [10]. (b) Calculated dependence (solid curves) of the electric strength E^* of rock salt on the duration of the leading edge of a pulse in a uniform field for various values of the distance between the electrodes: (*1*) 0.3 and (*2*) 0.15 mm. The displayed experimental data were borrowed from [8].

led to their burning, which resulted in the formation of a through channel. The experimental data in Fig. 2 (points) correspond to two distances between the electrodes used, 0.01 and 0.03 cm. One can see that, for $t_0 \leq$ 1.5 μs, the electric strength of the material increases

Parameter values used in calculating the time dependence of a pulsed breakdown

Material	τ, μs
Rock salt	0.025
Porcelain	0.02
Ammonium perchlorate	0.33

with decreasing duration of the leading edge of a voltage pulse (with increasing rate of growth of the voltage in the sample). For $t_0 \geq 1.5$ μs, the breakdown voltage becomes virtually independent of t_0.

In the case being considered, the electric-breakdown criterion corresponding to the structural–time approach [1, 2] can be represented in the form

$$\frac{1}{\tau} \int_{t-\tau}^{t} E(t')dt' \leq E_c, \qquad (4)$$

where E_c is the static electric strength of a material (it may be dependent on the distance between the electrodes) and τ is the incubation time of the electric breakdown of a material (it is determined by the kinetics of electron multiplication in the electric discharge).

We assume, for a first approximation, that the breakdown occurs at the leading edge of the voltage pulse and that the voltage in the sample used grows linearly. The electric field is uniform. The incident voltage pulse then has the form

$$U(t) = \frac{U^*}{t_0}[tH(t) - (t-t_0)H(t-t_0)].$$

Since the electric field is uniform, we have

$$E(t) = \frac{U^*}{t_0 d}[tH(t) - (t-t_0)H(t-t_0)], \qquad (5)$$

where U^* is the amplitude of the pulse, t_0 is the duration of its leading edge, d is the distance between the electrodes, and $H(t)$ is a Heaviside step function.

Substituting (5) into (4) and considering that the breakdown occurs at the leading edge of the pulse, we find that, for $t_0 \leq \tau$, condition (4) assumes the form

$$\frac{E^*}{\tau} \int_0^{t_0} t' dt' \leq E_c,$$

and that, for $t_0 \geq \tau$, the condition is

$$\frac{E^*}{\tau} \int_{t_0-\tau}^{t_0} t' dt' \leq E_c,$$

where $E^* = U^*/d$.

Since the breakdown time is determined by the instant at which the condition in (4) reduces to a strict equality, we obtain the following dependences of the electric strength on the duration of the leading edge of

the pulse:

$$
\begin{cases}
E^* = \dfrac{E_c}{1 - \dfrac{\tau}{2t_0}}, & t_0 \geq \tau \\[3mm]
E^* = \dfrac{2E_c\tau}{t_0}, & t_0 \leq \tau.
\end{cases} \tag{6}
$$

For some materials, the time dependences of the electric strength are shown in Figs. 2 and 3 (solid curves) according to a calculation based on (6). The parameter values used in this calculation are given in the table. In the displayed dependences, the instant at which the breakdown field begins increasing is completely determined by τ. In [9], it is indicated that this time is virtually independent of the distance between the electrodes. This also follows from the results that were obtained from our calculations and which are represented by the solid curves in Figs. 2 and 3. In the cases considered above, the incubation time can therefore be considered as a characteristic of a material.

Thus, it has been shown that experimental data on a pulsed breakdown are well described within our structural–time approach. The above examples of various physical processes indicate that it is of paramount importance to study incubation processes that prepare abrupt structural changes (fracture and phase transitions) in continuum media subjected to the effect of intense pulsed forces. The results presented here demonstrate that the structural–time approach, which makes it possible to describe adequately both the dynamical fracture of solids and the pulsed breakdown of solid dielectrics, is quite universal.

REFERENCES

1. Yu. V. Petrov, Dokl. Akad. Nauk SSSR **321**, 66 (1991) [Sov. Phys. Dokl. **36**, 802 (1991)].
2. N. Morozov and Y. Petrov, *Dynamics of Fracture* (Springer-Verlag, Berlin, 2000).
3. N. A. Zlatin, S. M. Mochalov, G. S. Pugachev, and A. M. Bragov, Fiz. Tverd. Tela (Leningrad) **16**, 1752 (1974) [Sov. Phys. Solid State **16**, 1137 (1974)].
4. P. Glebovskiĭ, in *Proceedings of the International Conference "Lomonosov-2001," Moscow, 2001* (Mosk. Gos. Univ., Moscow, 2001), pp. 246–247.
5. K. Ravi-Chandar and W. G. Knauss, Int. J. Fract. **25**, 247 (1984).
6. J. F. Kalthoff, Eng. Fract. Mech. **23**, 289 (1986).
7. J. W. Dally and D. B. Barker, Exp. Mech. **28**, 298 (1988).
8. A. A. Vorob'ev and G. A. Vorob'ev, *Electrical Breakdown and Fracture of Solid Insulators* (Vysshaya Shkola, Moscow, 1966) [in Russian].
9. I. G. Khaneft and A. V. Khaneft, Zh. Tekh. Fiz. **70** (4), 42 (2000) [Tech. Phys. **45**, 423 (2000)].
10. G. I. Skanavi, *Physics of Insulators (Strong-Field Region)* (Fizmatgiz, Moscow, 1958) [in Russian].

Translated by A. Isaakyan

<center>SOLIDS</center>

On One Supramolecular Mechanism of the Nonlinear Viscoelasticity of Oriented Polymers

A. M. Stalevich* and B. M. Ginzburg**

* State University of Technology and Design,
Bol'shaya Morskaya ul. 18, St. Petersburg, 191186 Russia

e-mail: stal@sutd.ru

** Institute of Problems in Machine Sciences, Russian Academy of Sciences,
Vasil'evskiĭ Ostrov, Bol'shoĭ pr. 61, St. Petersburg, 199178 Russia

Received February 27, 2004

Abstract—A phenomenological model of the viscoelasticity of highly oriented polymer systems is developed based on the results of studying the relaxation of such systems (mainly, fibers of polyethylene terephthalate, polyamide-6, polyvinyl alcohol, and other polymers) in the loaded state. The effect of an applied load on their relaxation spectra agrees qualitatively with the deformation behavior of crystal-like bandles that are present in amorphous intercrystalline layers of the fibrillar supramolecular structure. © 2004 MAIK "Nauka/Interperiodica".

The behavior of highly oriented polymer fibers under the action of an applied load is important for various engineering fields where such fibers or fiber-based composite materials are used. In this work, we study the mechanical properties of highly oriented fibers of polyamide-6, polyvinyl alcohol, polyethylene terephthalate, and some other polymers. Based on the experimental study of some processes characteristic of the engineering applications of polymer fibers (creep, uniaxial tension at various rates in the range of small strains, strain relaxation during complete or partial unloading in the range of nondestructive mechanical stresses, etc.), we proposed a phenomenological model for the viscoelasticity of these fibers [1–5]; it is described by the equation

$$\sigma_t = E_0\varepsilon_t + \int_{-\infty}^{\ln t} \varepsilon_{t-s}(\partial E_{\varepsilon s}/\partial(\ln s))d(\ln s), \qquad (1)$$

where σ_t and ε_t are the stress and strain at the time t; t is the duration of deformation; θ is the current time in the range from zero to t; $s = t - \theta$ is the period corresponding to the transition from the final strain ε_t to its current value of $\varepsilon_\theta = \varepsilon_{t-s}$ (Fig. 1); $E_{\varepsilon s}$ is the relaxation modulus depending on time and (as a parameter) strain; and $\partial E_{\varepsilon s}/\partial(\ln s)$ is the derivative of the relaxation modulus with respect to time, which can be interpreted as the relaxation-time distribution of relaxing particles to a first approximation [3].

For one of the simplest experiments (the determination of relaxation at $\varepsilon = $ const), the relaxation modulus can be calculated by Eq. (1):

$$E_{\varepsilon t} = \sigma_t/\varepsilon = E_0 - (E_0 - E_\infty)\varphi_{\varepsilon t}, \qquad (2)$$

where E_0 is the initial quasi-elastic value of the modulus, E_∞ is the quasi-equilibrium value of the modulus, and $\varphi_{\varepsilon t}$ is the normalized function of the time t that depends on the parameter ε and varies from zero at $t \ll \tau$ to unity at $t \gg \tau$ (where τ is the real relaxation time).

The experiment showed that, for the approximation of relaxation, it is convenient to use the probability integral

$$\varphi_{\varepsilon t} = (2\pi)^{-0.5} \int_{-\infty}^{V_{\varepsilon t}} \exp(-0.5z^2)dz, \qquad (3)$$

as $\varphi_{\varepsilon t}$. Here,

$$V_{\varepsilon t} = a_{n\varepsilon}^{-1}\ln t/\tau_\varepsilon \qquad (4)$$
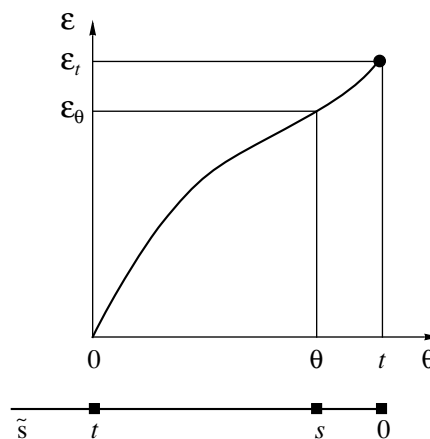
Fig. 1. Schematic diagram for integration when stresses are calculated by Eq. (1).

is the argument–functional that contains a constant $a_{n\varepsilon}$ and the average statistical relaxation time τ_ε, which depends on the strain as a parameter.

The above formulas indicate that, in its normalized form, the relaxation kernel under integral (1) is the Gaussian distribution

$$\partial\varphi_{\varepsilon s}/\partial(\ln s) = (2\pi)^{-0.5} a_{n\varepsilon}^{-1} \exp(-0.5 V_{\varepsilon s}^2), \qquad (5)$$

which can be interpreted, in a first approximation, as the normalized normal distribution of relaxing particles over the intrinsic relaxation times [13].

With allowance made for the relaxation (1), the relation between the stress and strain varying with time can also be written in the form [3]

$$\varepsilon_t = E_0^{-1}\sigma_t + \int\limits_{-\infty}^{\ln t} \sigma_{t-s}(\partial D_{\sigma s}/\partial(\ln s))d(\ln s) \qquad (6)$$

under the condition

$$E_0^{-1}E_{\varepsilon t} + \int\limits_{-\infty}^{\ln t} E_{\varepsilon;\,t-s}(\partial D_{\sigma s}/\partial(\ln s))d(\ln s) = 1, \qquad (7)$$

where $D_{\sigma t} = \varepsilon_{\sigma t}/\sigma$ is the compliance during simple creep $\varepsilon_{\sigma t}$ at $\sigma = \text{const}$.

The compliance measured is approximated similarly to the relaxation modulus by Eqs. (2)–(5). In this case, an analog of the average statistical relaxation time is the average statistical delay time (delay with respect to the quasi-equilibrium state) [3].

A typical example of the family of the time dependences of the relaxation modulus for polyfilament fibers of polyethylene terephthalate is shown in Fig. 2, and the corresponding strain dependence of the average statistical relaxation time and the stress dependence of the average statistical delay time are shown in Fig. 3. The descending character of these dependences means that, as the strain increases, both the relaxation and delay spectra shift to shorter times. This shift is a kind of nonlinearity of the viscoelastic properties that is caused by the activating effect of an applied mechanical stress on the relaxation [1, 2]. This shift can also be interpreted as the manifestation of a peculiar strain–time or force–time analogy [1–3]. Dependences that are similar to those shown in Figs. 2 and 3 are observed for highly oriented films and monofibers of polyethylene terephthalate and monofibers and polyfilament fibers of polyamide-6, polyvinyl alcohol, and polyacrylonitrile. It should be noted that, from the structural standpoint, polyacrylonitrile is not a typical amorphous–crystalline polymer: its ordered regions are characterized by ordering only across the axes of macromolecules (hexagonal packing of randomly rotating molecular cylinders [6]). However, the results of studying mechanical relaxation suggest that the supramolecular structure of highly oriented polyacrylonitrile fibers is similar to the structure of amorphous–crystalline polymers; that is, more and
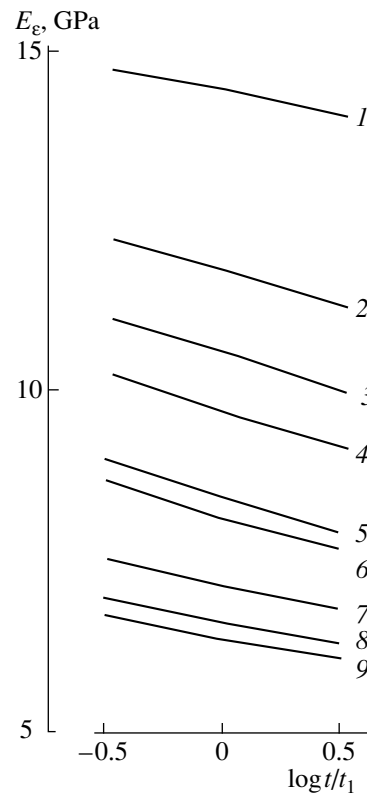
Fig. 2. Time dependences of the relaxation modulus of polyfilament fibers of polyethylene terephthalate at relative strain $\varepsilon = $ (*1*) 0.75, (*2*) 1.0, (*3*) 1.25, (*4*) 1.5, (*5*) 1.75, (*6*) 2.0, (*7*) 2.5, (*8*) 3.0, and (*9*) 3.5%; $t_1 = 1$ min.

less ordered regions alternate with each other along the fiber axis.

To reveal the most likely microscopic mechanism of the effect of an applied stress on the relaxation spectra, it is necessary to analyze the concepts of the supramolecular structure of highly oriented polymers. The intensity of the measured macroscopic relaxation of an applied stress is specified by the number of segment-molecular rearrangements that occur mainly in amorphous intercrystalline layers. If the number and volume of these interlayers were unchanged, no substantial deviations from linear viscoelasticity would be observed. In other words, Fig. 3 would show a horizontal straight line, i.e., a constant relaxation spectrum. The observed variation in the spectrum with strain (Fig. 3) is most likely due to the reversible increase in the weight fraction of the molecules that are in the amorphous state; these molecules have shorter relaxation times as compared to the molecules located in more ordered regions—crystallites and crystal-like bandles.

The concepts of crystal-like bandles in amorphous interlayers in amorphous–crystalline polymers (Fig. 4) were developed earlier [7–9]. In [7], the molecular-weight distribution of the products of etching of polyethylene films with fuming nitric acid was determined.
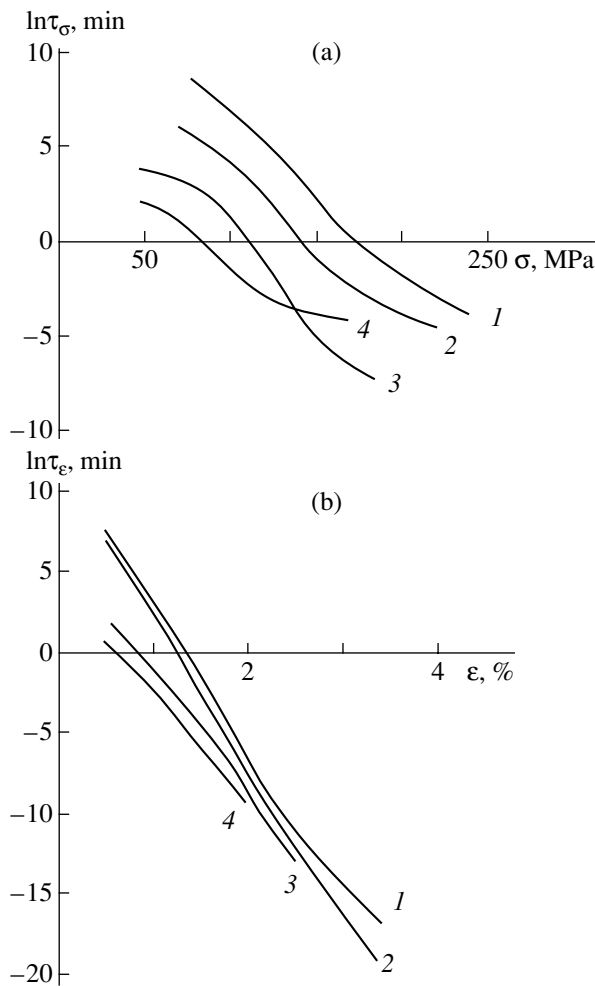
Fig. 3. Dependences of (a) the average statistical delay time on stress and (b) the average statistical relaxation time on strain for polyfilament fibers of polyethylene terephthalate at various temperatures: (*1*) 20, (*2*) 40, (*3*) 60, and (*4*) 80°C.
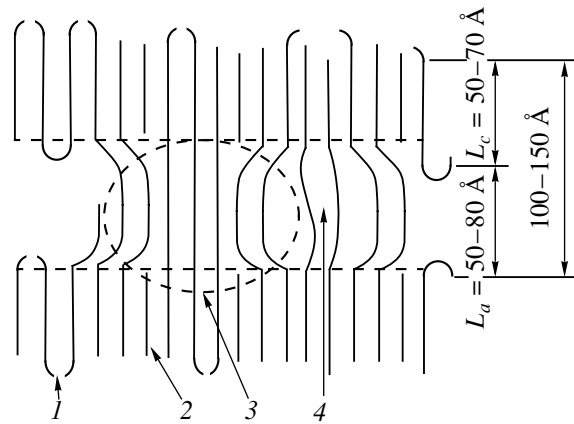

Fig. 4. Model of the supramolecular structure of highly oriented amorphous–crystalline polymers: (*1*) regions of folding macromolecules, which become destroyed when treated by fuming nitric acid [7]; (*2*) macromolecules in crystallites; (*3*) region of a dense crystal-like bandle in an amorphous interlayer; and (*4*) loose amorphous interlayers. The values of the long periods, crystallite sizes, and amorphous-interlayer sizes for low-density polyethylene data are taken from [7].

The acid destroyed macromolecules only in amorphous regions and in the places of packing of macromolecules at crystallite boundaries. The molecular-weight distribution was found to be bimodal: one distribution peak corresponded to the crystallite size along the chain and the other corresponded to the same size plus the length of the long period. Based on this finding, it was concluded that there exist dense crystal-like bandles in the amorphous regions, which are stable to the destroying action of the acid.

Tuĭchiev *et al.* [8] studied the effect of iodine sorption by fibers of polyvinyl alcohol and concluded that (i) molecular bandles with a high density (close to the crystal density) exist in amorphous regions and (ii) iodine cannot penetrate these bandles. In [9], a conclusion about the presence of high-density regions in amorphous interlayers was drawn from the changes in the intensities of highly oriented films of various polymers.

Schultz [10] noted that the absolute value of the integrated intensity of small-angle X-ray scattering from oriented fibers made of polyethylene terephthalate is significantly smaller than that predicted by a two-phase model for fibrils having long periods. In principle, this fact can be explained by the presence of interfibrila layers or defects in crystallites; however, in the context of the concepts developed in this work, this fact can also be explained by the presence of crystal-like bandles. As applied to polyethylene-terephthalate fibers, these concepts were confirmed in a number of works. For example, using NMR and the methods of small- and wide-angle X-ray diffraction, Biangardi and Zachmann [10] showed that, apart from a mobile amorphous component, polyethylene terephthalate fibers also contain a rigid, low-mobility amorphous component that consists of straightened highly oriented chain segments. In another work in the selection of articles [10], the amorphous halo in wide-angle X-ray diffraction patterns taken from polyethylene-terephthalate fibers was shown to correspond to a superposition of randomly and highly oriented components. Moreover, analyzing such data, the authors concluded that a third phase ("oriented mesophase") exists in these fibers. All these results agree with the concepts of crystal-like bandles.

Under a load or upon heating, the bandles decompose reversibly and the weight fraction of macromolecule segments in amorphous interlayers increases. Thus, the variation in the relaxation spectrum as a kind of the nonlinearity of viscoelastic properties of a highly oriented polymer can be explained by the reversible decomposition of the bandles located in amorphous interlayers. On the whole, this interpretation of nonlin-
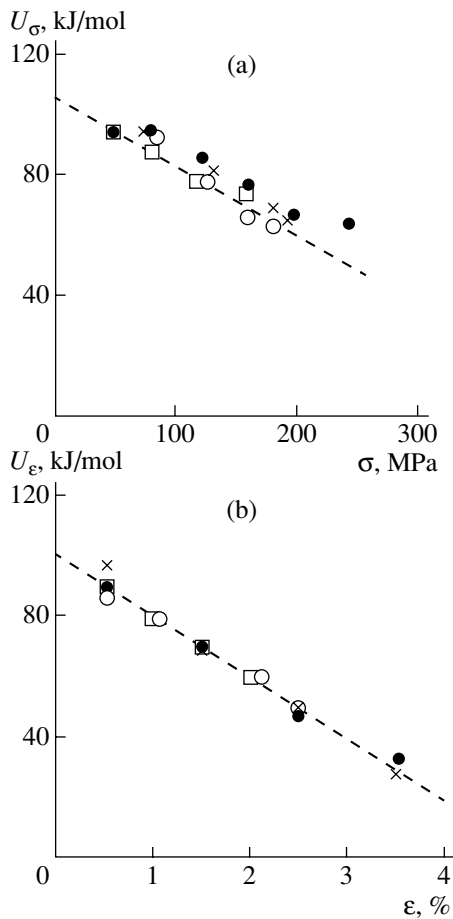
Fig. 5. Dependences of (a) the creep activation energy on stress and (b) the stress-relaxation activation energy on strain for polyfilament fibers of polyethylene terephthalate at various temperatures: (1) 20, (2) 40, (3) 60, and (4) 80°C.
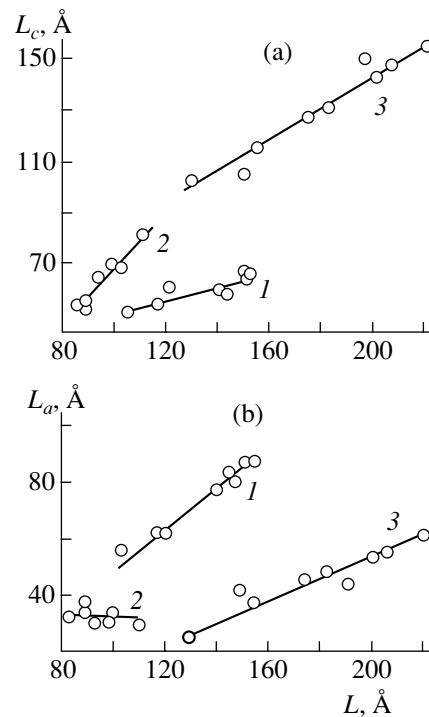

Fig. 6. Relations between the long period L and (a) the longitudinal crystallite size L_c and (b) the longitudinal amorphous-zone size L_a for (1) polyethylene terephthalate, (2) polyamide-6, and (3) polyethylene [11].

early inherited viscoelasticity is consistent with the results of other studies of the mechanical behavior of highly oriented fibers made of amorphous–crystalline polymers, specifically, recovery deformation processes that occur after complete or partial unloading [4] and after high-rate tension [5]. In addition, the specific role of the bandles described above is at least consistent with the deformation behavior of oriented amorphous–crystalline polymers, namely, the observed delay of recovery deformation processes with respect to the processes predicted by Eq. (1) [4], the dependence of the measured elastic modulus on the strain rate [5], and the possibility of taking into account the relaxation contributions at high strain rates [5].

The leading role of the bandle decomposition is also confirmed by the estimation of the potential barriers for macromolecular mobility, which controls the processes of creep and relaxation. The estimations were performed using the Arrhenius–Boltzmann formula

$$U_\sigma = RT \ln \tau_\sigma / \tau_0 \quad \text{or} \quad U_\varepsilon = RT \ln \tau_\varepsilon / \tau_0, \quad (8)$$

where τ_0 is the average statistical delay time determined from a family of the compliance curves measured under conditions of creep [3], τ_ε is the average statistical relaxation time determined from a family of the time dependences of the relaxation modulus [3], τ_0 is a constant, T is the Kelvin temperature, and R is the universal gas constant.

As can be seen from Fig. 5, the stress and strain dependences of the potential barriers calculated by Eq. (8) can be considered linear:

$$U_\sigma = U_0 - \gamma \sigma, \quad U_\varepsilon = U_0 - \gamma E_\tau \varepsilon, \quad (9)$$

where $E_\tau = 0.5(E_0 + E_\infty)$ is the average relaxation modulus.

At $\tau_0 = 10^{-13}$ s, we have the following numerical values of the energy constants: $U_0 = 100$ kJ/mol and $\gamma \approx 0.4$ nm^3. These values correspond to the segment mobility on the molecular scale of a structure, which does not contradict the assumed presence of crystal-like bandles in intercrystallite amorphous layers.

Let us estimate the size of elementary carriers of relaxation processes. The long periods in highly oriented amorphous–crystalline polymers with soft chains usually range from 10 to 20 nm, reaching 30–50 nm rather rarely at sufficiently high treatment temperatures. Figure 6 shows the most typical relations between the large-period size and the longitudinal sizes

of crystallites and amorphous regions for three polymers [11]; the length of amorphous interlayers in polyethylene terephthalate fibers varies from 5 to 8 nm. For quantitative estimations, we take their average value (6.5 nm). Then, the cross sectional area of the carrier is ~0.4/6.5 $\approx$ 0.06 nm^2, and it is ~0.25 nm across, which agrees in order of magnitude with the intermolecular distances in the crystal lattice and amorphous interlayers. Thus, the elementary carrier of the relaxation processes is likely to be a 5–8 nm segment of a macromolecule, which seems reasonable from the physical viewpoint. The number of monomeric units that are cooperatively involved in the relaxation processes depends on the length of the projection of a unit onto the texture axis. In particular, for polyethylene terephthalate, this length (lattice parameter c) is ~1.075 nm [12]. Then, a segment mobile element contains five to seven monomeric units.

We now analyze the activation energy obtained. One monomeric unit of polyethylene terephthalate has 24 atoms along the chain of a macromolecule (with allowance for the fact that each benzene ring has 6 atoms). Usually, the energy of intermolecular van der Waals interaction is 0.1–1 kJ/mol [13]. Then, no less than 100 atomic contacts (or no less than 4–5 monomeric units) take part simultaneously in the relaxation motion, which agrees satisfactorily with the size of the cooperative element estimated from the activation volume γ.

The estimates of the kinetic characteristics agree with the assumption that the reversible decomposition of the crystal-like bandles plays a leading role in the micromechanism of the observed effect of macrodeformation on the relaxation spectra.

REFERENCES

1. A. M. Stalevich, Probl. Prochn., No. 12, 95 (1981).
2. A. M. Stalevich, Izv. Vyssh. Uchebn. Zaved., Tekhnol. Tekstil. Prom-sti, No. 3, 95 (1980).
3. A. M. Stalevich, *Straining of Directed Polymer* (S.-Peterb. Gos. Univ., St. Petersburg, 2002) [in Russian].
4. A. M. Stalevich and L. E. Root, Probl. Prochn., No. 1, 43 (1988).
5. A. M. Stalevich, K. V. Sudarev, and Z. F. Stalevich, Probl. Prochn., No. 4, 86 (1986).
6. B. K. Vainshtein, *Diffraction of X-rays by Chain Molecules* (Izd. Akad. Nauk SSSR, Moscow, 1963; Elsevier, Amsterdam, 1966).
7. G. Meinel and A. Peterlin, J. Polym. Sci., Part A-2 **6**, 587 (1968).
8. Sh. Tuĭchiev, N. Sultanov, B. M. Ginzburg, and S. Ya. Frenkel', Vysokomol. Soedin., Ser. A **12**, 2025 (1970).
9. B. M. Ginzburg and N. Sultanov, Vysokomol. Soedin., Ser. A **43**, 1140 (2001).
10. J. M. Schultz, in *Oriented Polymer Materials*, Ed. by S. Fakirov (Hüthig and Wepf Verlag Zug, Heidelberg, 1996), pp. 361–393.
11. W. O. Statton, J. Polym. Sci. **41**, 143 (1959).
12. B. Wunderlich, in *Physics of Macromolecules*, Vol. 1: *Crystal Structure, Morphology, Defects* (Academic, New York, 1973; Mir, Moscow, 1976).
13. V. G. Dashevskiĭ and A. I. Kitaĭgorodskiĭ, in *Dictionary of Physics*, Ed. by A. M. Prokhorov (Sov. Éntsiklopediya, Moscow, 1984), p. 400 [in Russian].

Translated by K. Shakhlevich

================================ **SOLIDS** ================================

Plastic Deformation and Quasi-Periodic Vibrations in a Tribological System

V. E. Rubtsov and A. V. Kolubaev

Institute of Strength Physics and Materials Science, Siberian Division, Russian Academy of Sciences,
Akademicheskiĭ pr. 2/1, Tomsk, 634021 Russia
e-mail: root@ispms.tomsk.ru
Received March 11, 2004

Abstract—A one-dimensional macroscopic model is used to analyze the plastic deformation of materials without coating and with a plastic hardening coating or a plastic nonhardening coating at friction. The calculations show that mechanical vibrations can be excited in a tribological system and that their frequency decreases sharply when going from elastic to plastic deformation. One of the causes of the development of plastic deformation in the surface layer and in the sublayer of the material under a hard coating is found to be a decrease in the elastic properties of the material because of frictional heating. An intense plastic shear in the material under the hard coating can cause its failure due to incompatible strains of the coating and the base. © 2004 MAIK "Nauka/Interperiodica".

INTRODUCTION

Friction is a widely occurring physicomechanical process that takes place when surfaces are in contact. For example, it occurs when we try to keep our balance on a slippery road, to repair our old car using wonder-working additives for lubricants, to choose the material for parts of a designed mechanism, and so on. Most processes that are caused by the interaction and motion of system elements obey fundamental laws, which can be used to predict the behavior of the system; however, friction induces an uncertainty that depends on many factors connected by relations that can be determined only experimentally. To take into account all the relations characterizing the behavior of friction pairs, one has to use the achievements of surface physics, the mechanics of contact phenomena, strength physics, plasticity physics, mechanochemistry, electrochemistry, and many other sciences. Not all the explicit and implicit factors can be taken into account. Therefore, the most realistic approach to understanding the processes of friction and wear is based on studying various aspects of this phenomenon when basic parameters that determine the state of a certain tribological pair can be separated.

From the physical standpoint, the most important problem is the problem of degradation of the surface layer of a material; its solution should be based on understanding the leading role of plastic deformation at friction and the related factors (mainly dynamic processes). The dynamic character of friction manifests itself in macroscopic mechanical vibrations that appear in a friction system upon sliding. It was found in [1] that one of the causes of the formation of a deformed near-surface layer at friction is mechanical and thermal excitations that propagate deep into the medium and are caused by the transition of a friction system into a self-vibrating state with a large amplitude of mechanical vibrations. In this case, temperature oscillations occur at the friction surface and in a subsurface layer; as a result, thermal cyclic creep and phase and structural transformations may develop in the material.

Interest in dynamic processes has increased in the past few years, and they are extensively studied both experimentally and theoretically. The purpose is to study the causes of vibrations in a tribological system, the conditions of stability of various vibration modes, and the critical values of parameters and their combinations that change the system into a specific friction mode.

During contact interaction, especially in the case of dry friction, quasi-periodic or periodic changes in the velocity of the relative motion of the surfaces are observed experimentally. On the macroscale, this behavior manifests itself in the form of relaxation or quasi-harmonic friction self-sustained vibrations. There are several approaches to explaining the causes of vibrations in a friction system. In all of them, the development of vibrations is related to changes in the friction coefficient in frictional interaction. In one of the approaches [2], the occurrence of vibrations is explained by the descending nonlinear dependence of the friction coefficient on the sliding velocity. In another model [3], the friction force is considered velocity-independent and self-sustained friction vibrations are assumed to be developed because of a positive difference between the static friction force and the kinetic friction force. In the third approach [4], both a positive difference between the static and kinetic friction forces and a change in the friction force with

changing velocity are taken into account. It is shown experimentally that rheological phenomena in the contact region play a significant role in the appearance of friction self-sustained vibrations [5]. In particular, the appearance of friction self-sustained vibrations is related to temperature-induced changes in the friction characteristics [6]. This model assumes that, in the stage of relative sliding in a vibration cycle, the heating of friction surfaces results in a decrease in the friction coefficient and that, at the stage of relative rest, the surfaces cool down and the friction coefficient increases. Indeed, a decrease in the friction coefficient with increasing temperature is observed experimentally.

The dynamic character of friction manifests itself in not only the macro- but also the microscale. On the microscale, a change in the friction force is caused by stick-slip motion. In this case, the relative motion of the two surfaces is discontinuous and consists of sequential stick–slip cycles [7]. The mechanism of this oscillatory motion of surfaces in friction has not yet been understood; however, most researchers believe that it is also related to the dependence of the friction coefficient on the sliding velocity [8].

MODEL OF A TRIBOLOGICAL CONTACT

In this work, we analyze the dynamic behavior of a tribological contact, which specifies the viscoelastic deformation of the surface layers at friction, using a one-dimensional macroscopic model. The strain of the subsurface layers and the propagation of strain into the bulk of a material are determined with allowance for changes in its properties due to the strain hardening and frictional heating.

From the macroscopic standpoint, the structure and properties of the material formed within a contact spot are virtually uniform in the direction parallel to the friction surface and vary only with the distance from this surface. Therefore, to study shear strain in the surface

layer in the simplest case, we can restrict ourselves to one-dimensional approximation. The model to be developed is macroscopic not from the standpoint of a spatial scale: this model does not take into account the structure and deformation mechanism of the medium under study and deals only with its macroscopic characteristics.

Frictional interaction was taken into account within a contact spot, which is represented as a surface microasperity consisting of a number of layers having different physicomechanical properties in the general case (Fig. 1). The response of the system to an external action is determined by the properties of the layers forming the microasperity and the laws of interaction between them. This approach allows us to describe and to arbitrarily specify a property gradient in the microasperity and to design this microasperity from several materials with different properties. The layers are assumed to be absolutely rigid and to have the same length L (along the surface) and thickness h_1. Compression induced by a contact (normal) pressure P_{cont} is not taken into account explicitly in the model. The normal stress is assumed to be constant over the whole height of the microasperity and equal to the contact pressure. We take into account the normal stress when calculating the plasticity criterion.

A shear that can be induced by a friction force in the microasperity is realized in the model by shifting the layers with respect to each other parallel to the friction surface. The elementary shear carrier is a pair of layers. Each layer is assumed to interact only with the two nearest neighbors. To find the shear stress operating in a pair of layers, we use a viscoelastic response function, which is similar to the function used in [9]. It has the following parameters: elastic and "plastic" moduli, the maximum elastic strain, and an equilibrium shear. The parameters of the response function are varied with the deformation history and the temperature using a special procedure.

The material of each layer is characterized by a density ρ, specific heat c, thermal conductivity λ, shear modulus G, plastic shear modulus G_{pl}, and yield strength upon uniaxial tension Y. Moreover, each layer has a shift x and a translational velocity V, which are found by integrating a set of classical equations of motion for the whole system of the layers expressed in terms of stresses (1)–(3) with boundary condition (4):

$$\rho^1 h \frac{dV_n^1}{dt} = \tau_n^{fr} + \tau_n^{1,2}, \tag{1}$$

$$\rho^i h \frac{dV_n^i}{dt} = \tau_n^{i,i-1} + \tau_n^{i,i+1}, \tag{2}$$

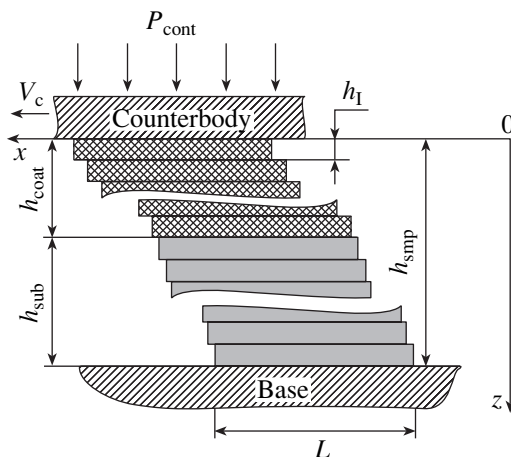$$\frac{dx_n^i}{dt} = V_n^i, \tag{3}$$

Fig. 1. System being simulated.

$$\frac{dx_k^i}{dt} = 0, \tag{4}$$

where i is the number of a layer ($i = 1, 2, \ldots, k-1$; index 1 is referred to the upper layer; and k is the number of layers), $\tau_n^{\text{fr}} = P\mu_n$ is the friction force-induced shear stress at the surface, and μ_n is the friction coefficient at the nth time step.

Equation (1) is written for the upper layer of the microasperity. The upper layer slides on a counterbody that moves at a constant velocity V_c. Boundary condition (4) is the condition of fixation of the lower layer on the substrate.

At the beginning of calculation, the properties of a pair of layers are taken to be the average properties of its elements. During simulation, the mechanical properties of the pair vary with temperature.

To find the temperature field in the microasperity, we solve a one-dimensional heat problem that is similar to that in [10]. The choice of boundary conditions strongly affects the solution of the heat problem and, hence, the behavior of the friction system. As a rule, we use the condition of ideal thermal insulation to calculate the temperature field in the microasperity using the one-dimensional approximation and to represent the region to be simulated in the form of a thin rod for the microasperity side that is opposite to the contact surface. In our case, the penetration depth of thermal perturbation exceeds the microasperity height because of long simulation times and taking into account the cyclic character of frictional interaction. Therefore, the boundary condition of ideal thermal insulation is invalid for the base of the microasperity. One of the methods to solve the problem is to use the boundary condition in the form

$$T(h_{\text{smp}}, t) = T_{\text{bulk}}(h_{\text{smp}}, t),$$

where $T_{\text{bulk}}(h_{\text{smp}}, t)$ is the average temperature at a depth $z = h_{\text{smp}}$.

We find $T_{\text{bulk}}(h_{\text{smp}}, t)$ from the solution of the problem of frictional heating of the sample for a larger depth with a larger spatial grid step, disregarding the discrete character of heat release at the surface during friction. To calculate this temperature, the intensity of heat release at the contact surface was determined from the nominal rather than contact pressure. Thus, the temperature of the microasperity base is related to the average temperature of the sample at a distance h_{smp} from the surface, which increases monotonically during friction. This is schematically shown in Fig. 2. The application of the above boundary condition restricts heating in the microasperity, which corresponds to intense heat removal to the neighboring unheated regions near the contact spot.

To calculate the viscoelastic strain, we use the deformation criterion of plasticity. The material in a pair of neighboring layers is thought to undergo viscoelastic
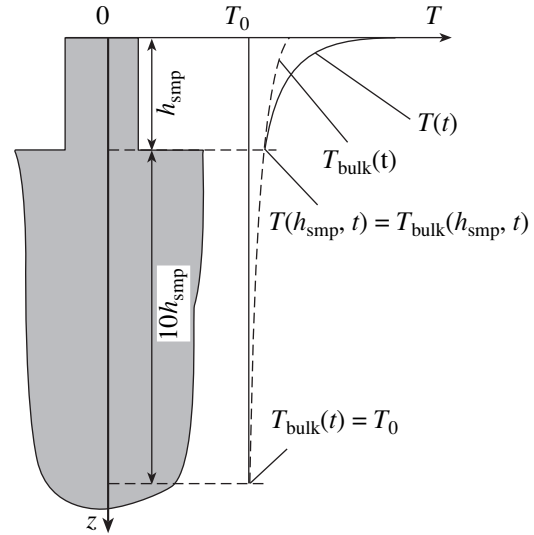
Fig. 2. Schematic diagram and the boundary conditions for solving the heat problem.

deformation when the strain at the nth time step exceeds the maximum elastic strain for the current temperature in this pair; i.e., when $\gamma_n^{i, i+1} > \gamma_{n\text{el}}^{i, i+1}(T_n^{i, i+1})$, where $T_n^{i, i+1}$ is the temperature in the pair of layers i and $i + 1$. The value of $\gamma_{n\text{el}}^{i, i+1}(T_n^{i, i+1})$ depends on the temperature and the deformation history of the pair of layers.

Before simulation, we calculated the reference values of the maximum elastic strain for each pair of layers over the entire temperature range $\gamma_{0\text{el}}^{i, i+1}(T^{i, i+1})$.

Although the model is one-dimensional, we found plasticity conditions using the Mises criterion for plane strain (after Taybor [11]). The application of this criterion is grounded on the fact that two stresses (normal and shear) are present in the microasperity. With this criterion, we can take into account the contribution of the normal stress in terms of the one-dimensional formulation of the problem. Based on the above considerations, we determine the maximum elastic strain $\gamma_{0\text{el}}^{i, i+1}(T^{i, i+1})$ at a given temperature from the expression

$$\gamma_{0\text{el}}^{i, i+1}(T^{i, i+1}) = \frac{\sqrt{(Y(T^{i, i+1}))^2 - P_{\text{cont}}^2}}{\sqrt{3}G^{i, i+1}(T^{i, i+1})}. \tag{5}$$

The value of $\gamma_{0\text{el}}^{i, i+1}(T^{i, i+1})$ is used in the response function to calculate the dynamic behavior of the system when the material in a pair of layers has not undergone plastic shear until a given time. The response function relates the shear strain to the shear stress for a pair of layers and can describe the viscoelastic strain for both hardening and nonhardening materials. The

parameters of the response-function may vary as a result of heating and due to plastic strain of the material.

CALCULATION OF THE INTERACTION OF A SINGLE MICROASPERITY DURING FRICTION

The microasperity consists of 150 1-μm-thick layers. When the microasperity has a coating, the corresponding number of its upper layers of thickness h_{coat} have properties other than those of the base material. The material of a hard coating is taken to be ideally elastic with a shear modulus that is twice as large as the shear modulus of the base material. The material of a plastic coating has a yield strength that is half that of the base material. Other properties of the coatings are identical to the properties of the base material, which are chosen to be similar to the properties of 12Kh18N10T austenitic steel.

Our purpose was to analyze the model material rather than simulate the evolution of a certain material at friction. Therefore, for simplicity, we assumed that the mechanical properties decrease linearly with increasing temperature. We also assumed that, at a temperature of 1300°C (which is close to the melting temperature), the shear modulus, the plastic shear modulus, and the yield strength are 10% of their values at 20°C.

An elementary event of interaction of the microasperity is simulated at counterbody sliding velocity $V_c = 1$ m/s, coefficient of sliding friction $\mu = 0.5$, contact pressure $P_{cont} = 162$ MPa, nominal pressure $P = 7.5$ MPa, and initial temperature $T_0 = 20$°C.

Before simulation, the sample is shifted from an initial state to an equilibrium position; that is, the layers are displaced with respect to each other so that the elastic shear stress balances the friction force-induced stress. This provides a zero velocity at the initial instant

and the absence of plastic deformation at the interface with the rigid base during the calculation. In so doing, we excluded the effect of fixation of the lower part of the sample on the simulation results.

The time of interaction of the microasperity with the counterbody t_{cont} is determined from the relation between the contact spot size L in the sliding direction and the counterbody velocity V_c: $t_{cont} = L/V_c$. Thus, when simulating, the contact spot size is specified by the duration of the interaction of the microasperity with the counterbody.

We calculated the following systems: an uncoated material, a material with a hard coating, a material with a plastic strain-hardening coating, and a material with a plastic nonhardening coating.

The simulation shows that, at the initial instant, the sample is elastically deformed under the action of the friction force; further sliding results in heat release at the friction surface and heating of the microasperity. This heating is accompanied by a decrease in the elastic properties (softening) of the surface-layer material. However, for a short contact time, the increase in the temperature is insignificant, and the degree of softening is insufficient for plastic deformation. Curve *1* in Fig. 3 shows the velocity field of the system at a certain time when the strain is still elastic. Plastic deformation is absent in this simulation stage, since the increase in the temperature does not cause substantial changes in the mechanical properties of the material. At this instant, the system exhibits quasi-periodic vibrations at a frequency of several megahertz. The vibrations result from different mechanical properties of the surface layers due to the temperature gradient. Moreover, the initial displacement of the sample, which balances the friction force, affects the vibrations.

Increasing the temperature to a certain critical value at which the friction-induced tangential stresses reach the yield stress causes the plastic shear of the surface layers. As the sample is heated, plastic deformation involves new surface layers. The dynamic response of the system is interesting in this case. Plastic deformation generates vibrations in the system whose frequency is about an order of magnitude lower than the frequency observed upon elastic deformation (Fig. 3, curve *2*). These vibrations are unstable because the plastic shear in the surface layer occurs stepwise rather than continuously. Every new plastic-deformation event causes generation of an elastic shear wave in the underlying material, which propagates deep into the bulk at a velocity corresponding to that of the transverse sound wave. Then, the elastic shear wave formed continues to move along the microasperity, reflecting periodically from the rigid base and the surface. The interference of all the waves induced by plastic shears gives the pattern shown in Fig. 3 (curve *2*). The vibration frequency turns out to be approximately equal to the tribosystem eigenfrequency.
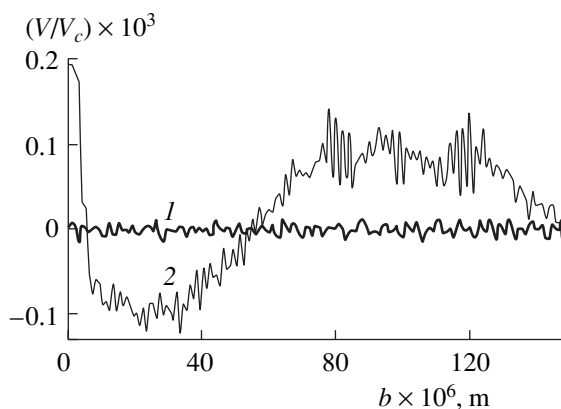

Fig. 3. Ratio of the layer velocity V in the sample to the counterbody velocity V_c (b is the distance from the friction surface): (*1*) before the beginning of plastic deformation and (*2*) at a plastic strain of $\approx 0.6\%$ at the surface.

Vibrations of this type were experimentally observed as the vibrations of a friction unit or as repetitive sound signals. The study of acoustic emission indicates that a decrease in the acoustic-radiation frequency is also related to the beginning of plastic deformation. It should be noted that, in a real tribological conjugation, the frequencies of such vibrations are several orders of magnitude lower than the calculated ones, since the sizes of real tribosystems are several orders of magnitude larger than those of the system simulated.

Another manifestation of an increase in the temperature of the surface layers is the nonlinear depth dependence of the yield strength of the material. It can be tracked in our calculations using the current value of the maximum elastic strain γ_{el}, which is proportional to the yield strength. Figure 4 shows the change in the ratio of the current maximum elastic strain to its initial value with increasing the distance from the surface. This behavior of γ_{el} is controlled by strain hardening of the material during plastic deformation. On the one hand, heated and plastically deformed layers become softened, and, on the other hand, the overlying layers become hardened. The minimum ratio $\gamma_{el}/\gamma_{0\,el}$ is specified by the viscoelastic properties of the material and the limiting temperature above which the friction force induces plastic deformation in the surface layers. To decrease the simulation time, we chose the initial temperature and normal pressure such that plastic deformation started at a very small increase in the temperature in the calculation. Therefore, the minimum is low. Figure 5 shows the shear of a relatively thick copper layer on a sample (block) as a result of a sharp increase in the load at friction.[1] One of us observed a similar effect of a metal shear on the brake block of a railway car: it was covered with metal layers removed from a railway wheel as a result of emergency braking. This phenomenon and the shear shown in Fig. 5 can be explained using the curves of Fig. 4. In the steady-state friction mode, the shear rate decreases gradually to zero with the depth due to a plastic flow. A sharp increase in the load on the friction surface (e.g., emergency braking) causes an increase in the stresses in the underlying layers, and they exceed the yield stress, which decreases upon heating during the previous friction. In this case, the velocity of all the surface layers can increase stepwise so that the softest layers shift with respect to the base material at a certain depth.

SIMULATION OF FRICTION IN COATED MATERIALS

The plastic deformation caused by a decrease in the mechanical properties of the material can lead to failure of hard wear-resistant coatings. The results of friction tests of samples with coatings produced by nitrogen ion-beam treatment are given in [12]. Layers 3–6 μm thick did not give a substantial increase in the wear

---
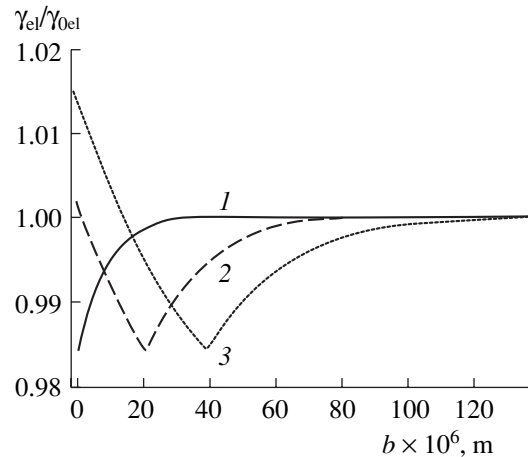
Fig. 4. Ratio of the current maximum elastic strain γ_{el} to its initial value in the sample $\gamma_{0\,el}$ at different times: (*1*) 1.6×10^{-5}, (*2*) 1×10^{-4}, and (*3*) 2×10^{-4} s.
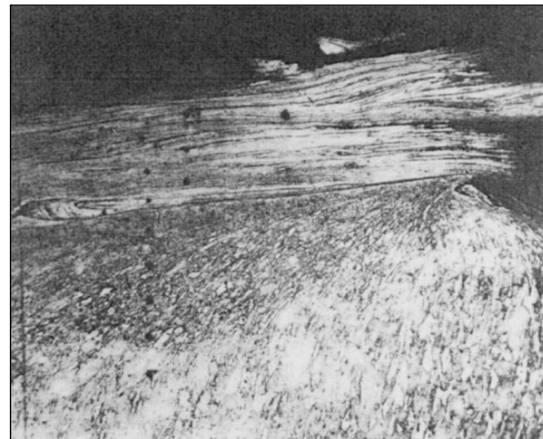
Fig. 5. Shear of the surface layers in copper induced by a sharp increase in the normal (contact) load at friction.

resistance of the steel surface under the conditions of dry-friction contact interaction. During the tests, the modified layer failed in the first meters of the sliding distance and the wear intensity reached the level of the unimplanted steel. Thicker hard coatings have a higher wear resistance.

Our calculations indicate that plastic deformation, which is mainly caused by the softening of the base material because of frictional heating, develops under a hard layer. If the time of frictional interaction in the contact spot exceeds the certain critical time it takes for the base material to be softened, an intense plastic shear occurs in the sublayer under the coating. This shear may result in the failure of the coating because of deformation incompatibility between the hard layer and the base material. If the coating thickness provides a significant decrease in the temperature and stress in the sublayer so that they do not cause a plastic flow of the base
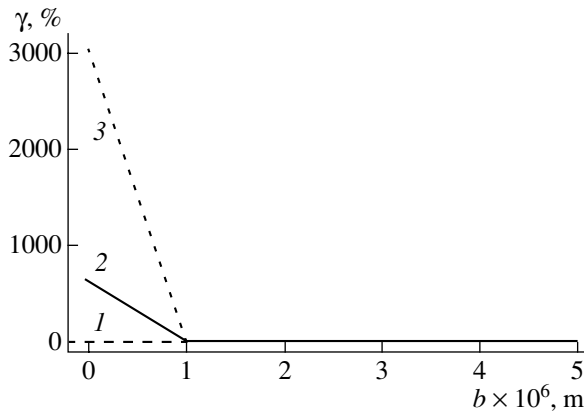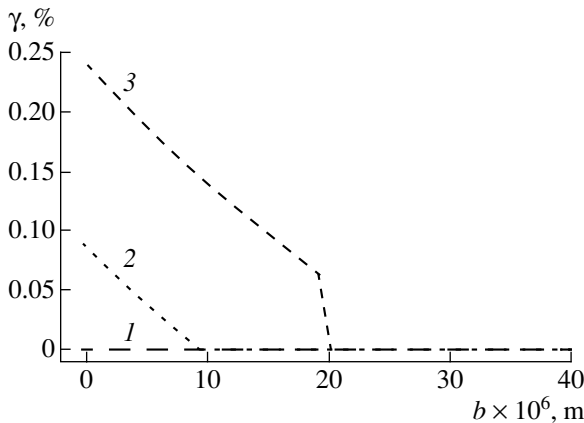
Fig. 6. Dependence of the shear plastic strain γ on the distance b from the friction surface in a sample with a 20-μm-thick nonhardening plastic coating at different times: (*1*) 1.6×10^{-5}, (*2*) 2.5×10^{-5}, and (*3*) 5×10^{-5} s.

Fig. 7. Dependence of the shear plastic strain γ on the distance b from the friction surface in a steel sample with a 20-μm-thick hardening plastic coating at different times: (*1*) 1.6×10^{-5}, (*2*) 5×10^{-5}, and (*3*) 1.5×10^{-4} s.

material, the coating does not fail. Otherwise, the plastic deformation under the hard coating would be incompatible with the elastic deformation of the coating, which causes cracking in the latter.

The friction simulation of a coating made of an ideally plastic material indicates that a plastic shear occurs only in the upper pair of elements irrespective of the coating thickness (Fig. 6). Until the beginning of plastic flow (Fig, 6, curve *1*), the temperature in the upper layer increases. As the temperature increases, the coating material becomes softened; when the flow stress becomes lower than the shear stress caused by the friction force, the upper layer is entrained by the counterbody and plastic shear develops. In time, the plastic shear becomes very high (Fig. 6, curves *2*, *3*). After the beginning of the plastic flow, the velocity of the upper layer is not constant. It oscillates with a certain amplitude about an average value, which is ≈70% of the

counterbody velocity. Correspondingly, the velocity of sliding between the microasperity and counterbody is significantly lower than its initial value. Therefore, the heat release at the surface decreases several times and the surface temperature decreases. It should be noted that the plastic shear does not terminate after the temperature decreases below the value at which softening of the coating begins. This behavior is related to the dynamic character of loading that occurs after the beginning of the plastic flow (see above). Figure 7 shows plastic-deformation curves at various times for a hardening plastic coating. The curves are similar to the behavior of the uncoated plastic material described in the previous section. Plastic deformation begins on the surface, and the whole coating becomes plastic in the course of time. The less plastic base remains elastic. The temperature in the sample increases monotonically.

CONCLUSIONS

(1) A comparison of the simulation results with experiment shows that, despite its simplicity, the model proposed can correctly describe the behavior of gradient materials during frictional interaction. It can describe the character of dynamic excitations that develop as a result of the interaction of contacting surfaces in the contact spots and are accompanied by high-frequency vibrations. As follows from the model calculations, one of the possible causes of the vibrations, which manifest themselves in real friction pairs in the form of macrovibrations and sound, can be a plastic shear of the surface layer in the contact spots.

(2) One of the main results is the fact that the thermal conditions in the surface layer are a key factor that determines the behavior of a tribosystem. Changes in the tribosystem parameters, such as the mechanical properties of the material, loading conditions, and the contact geometry, affect the plastic deformation of the surface layer both directly and indirectly because of a change in the thermal conditions in the contact zone. The same thermal processes are substantially responsible for the failure of hard coatings, since an incompatible deformation develops in the hard layer and the base material as a result of the softening of the base material due to frictional heating. This deformation leads to failure of the coating.

(3) The friction simulation of a material with a plastic coating shows that hardening and nonhardening coatings behave differently. In the case of a hardening material, plastic deformation develops throughout the coating. In the case of a nonhardening coating, plastic deformation is localized in a thin surface layer and reaches very high values (1000% or more) during the contact. After the beginning of intense plastic deformation in the plastic layer, the friction-surface temperature decreases substantially. Thus, the plastic nonhardening layer serves as a solid lubricant and decreases strongly frictional heating.

SOLID-STATE ELECTRONICS

Simulation and Improvement of the Selectivity of Crystal-Like Structures

E. A. Nelin

National Technical University of Ukraine Kiev Polytechnic Institute, Kiev, 03056 Ukraine
e-mail: ye_nelin@rtf.ntu-kpi.kiev.ua
Received September 10, 2003; in final form, April 29, 2004

Abstract—A model of equivalent reflectors proposed for crystal-like structure makes it possible to obtain the dispersion characteristic in explicit form. The similarity of the dispersion characteristics of waveguides and crystal-like structures is demonstrated. A fundamental solution is proposed for improving the selectivity of such structures. © *2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Artificial periodic structures similar to natural crystals have become objects of extensive study in recent years. These structures form the basis for various new (primarily nanoelectronic) integrated signal processing devices. Among such crystal-like structures (CSs), considerable advances have been made in the field of semiconducting superlattices (SLs) for electrons as de Broglie waves, photonic crystals (PtCs) for electromagnetic waves, and phonon crystals (PnCs) for elastic waves. In the frequency bands formed in crystals and CSs due to constructive interference of reflected waves (forbidden energy bands for electrons and forbidden gaps for photons and phonons), waves cannot propagate. Such a frequency filtration makes it possible to control the transmission of waves in CSs, which extends the potential of signal processing methods.

Simulation of crystals and CSs is reduced to solving the wave equation with periodic boundary conditions mainly using the matrix method [1] and finite-difference time-demain analysis [2]. The relatively simple models proposed recently [3, 4] are based on analysis of the Airy formulas for the coefficients of reflection and transmission of the optical layer and on the analogy between the dispersion relation for a PtC and a pseudoquantum particle, respectively, make it possible to establish important features in the spectral characteristics of PtCs.

In this study, we propose a general model of equivalent reflectors for CSs, which provides quite accurate (and even explicit) expressions for dispersion characteristics and permits a visual analysis and optimization of constructive solutions for CSs. This model serves as the basis of a method for improving the CS selectivity.

Principal features of CSs are manifested in one-dimensional structures. Let us consider the dispersion characteristics of such CSs.

DISPERSION CHARACTERISTICS OF CS: MODEL OF EQUIVALENT REFLECTORS

The dispersion characteristics of PtCs and PnCs formed by layers with alternating refractive indices and with alternating acoustic properties, respectively, for the normal incidence of a wave have the form [1, 5, 6]

$$\cos K\Lambda = \cos k_1 a \cos k_2 b - \frac{1+\chi^2}{2\chi}\sin k_1 a \sin k_2 b, \quad (1)$$

where K is the Bloch wave number; Λ is the period of the structure; a and b are the thicknesses of the layers; $k_{1,2}$ are the wave numbers corresponding to the layers; and $\chi = k_1/k_2 = n_1/n_2$ ($n_{1,2}$ are the refractive indices of the layer) for PtCs and $\chi = c_1 k_1/c_2 k_2$ ($c_{1,2}$ are the elastic moduli of the layers) for PnCs.

It should be noted that the dispersion characteristic for PnCs in recent publications [3, 7] contains an error due to the neglect of the elastic moduli.

The dispersion characteristic of a SL formed by alternating potential barriers and wells in the case of electron tunneling is defined as [8]

$$\cos K\Lambda = \cosh k_1 a \cos k_2 b - \frac{1-\chi^2}{2\chi}\sinh k_1 a \sin k_2 b, \quad (2)$$

where $k_1 = \sqrt{2m(V-E)}/\hbar$, $E < V$; $k_2 = \sqrt{2mE}/\hbar$, m is the effective electron mass, V is the height of potential barriers, E is the electron energy, $\hbar = h/2\pi$, h is the Planck constant; a is the barrier width, b is the well width, and $\chi = k_1/k_2$.

For different effective electron masses in the region of a barrier and a well (m_1 and m_2, respectively), we have $\chi = k_1 m_2/k_2 m_1$ in view of the difference in the boundary conditions for these regions [9].

If we use the general expression for k_1, $k_1 = \sqrt{2m(E-V)}/\hbar$, k_1 is imaginary for $E < V$ and formula (2) follows from Eq. (1) after the substitution of imagi-

nary k_1. Thus, Eq. (1) establishes the implicit dependence of the dispersion characteristic for all CSs considered here.

We will prove that quantity χ is equal to the direct or inverse ratio of impedances Z_1 and Z_2 of the layers. For optical waves, we have $\chi = n_1/n_2 = Z_2/Z_1 = r_Z^{-1}$. For elastic waves, $Z = v\rho$, where v is the velocity and ρ is the density of the medium. Since $v = \sqrt{c/\rho}$, we have $\chi = c_1 v_2/c_2 v_1 = Z_1/Z_2 = r_Z$. For electron waves [10, 11], $Z = \pm 2\sqrt{2(E - V)/m}$, where the plus and minus signs correspond to the positive and negative directions of wave propagation. For $E < V$, the impedance is imaginary, which corresponds to electron tunneling in a bounded medium (potential barrier); for $E > V$, the impedance is real, which corresponds to passage of electrons above the barrier. Since $Z = \pm 2\hbar k/m$, for electron waves we have $\chi = Z_1/Z_2 = r_Z$.

Thus, the first cofactor in the second term on the right-hand side of Eq. (1), as well as the coefficient of reflection from the interface between the layers with impedances Z_1 and Z_2, is determined by the ratio r_Z of impedances.

We transform Eq. (1) for real-valued k_1 and Z_1 by adding the term $\sin k_1 a \sin k_2 b$ to the right-hand side of Eq. (1) and subtracting it,

$$\cos K\Lambda = \cos kD - \frac{(r_Z - 1)^2}{2r_Z}\sin k_1 a \sin k_2 b, \quad (3)$$

where $D = n_1 a + n_2 b$, $k = \omega/c$; for optical waves, $n_{1,2}$ are the refractive indices of the layers and c is the velocity of light in vacuum; for elastic and electron waves, quantities $n_{1,2}$ are analogous to refractive indices $n_{1,2} = c/v_{1,2}$, $c = (v_1 + v_2)/2$.

Disregarding the second term on the right-hand side of Eq. (3) and the periodicity of function $\cos x$, we have $K\Lambda = kD$. In this case, the dispersion characteristic defined by expression $\omega = cK/\langle n \rangle$, where $\langle n \rangle = (n_1 a + n_2 b)/(a + b)$ is the averaged refractive index, corresponds to the case when reflections within the structure are disregarded. Thus, the second term on the right-hand side of Eq. (3) is associated with the effect of reflections.

In contrast to optical and electron waves, for elastic waves it is possible that $v_1 \neq v_2$ and $r_Z = 1$ in some cases (if $v_1/v_2 = \rho_2/\rho_1$). Reflections and, accordingly, the second term in Eq. (3) are absent, and a band diagram is not formed.

Let us see how this term is connected with the coefficients of reflection from the layers (see the inset to Fig. 1). The arrows denote the sources of reflection at the boundaries of the layers; $r = |r_Z - 1|/(r_Z + 1)$ is the modulus of the coefficient of reflection from the layer boundary. The direction of the arrows corresponds to $r_Z > 1$. We halve the sources and sum the halves, reduc-
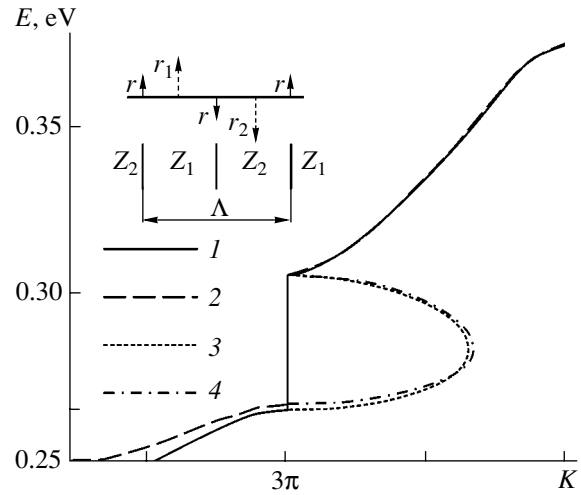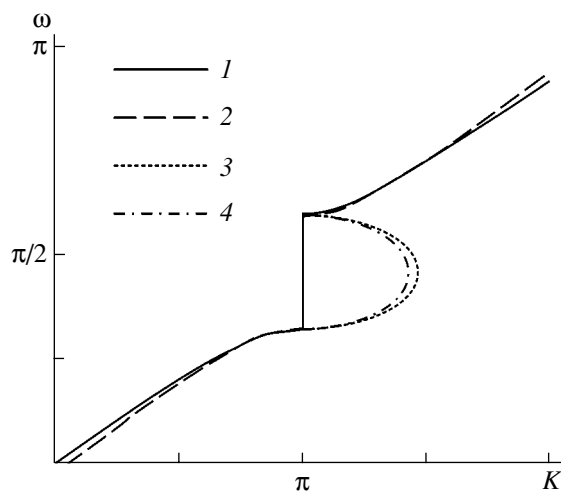

Fig. 1. Dispersion characteristics of a SL. Curves *1*, *2* and *3*, *4* represent the real and imaginary parts of K in accordance with relations (1), (5) and (1), (6). The real part of K is expressed in units of $1/\Lambda$ and the imaginary part, in arbitrary units.

ing then to the centers of the layers (sources $r_{1,2}$ in Fig. 1): $r_1 = 0.5r\exp(ik_1 a) - 0.5r\exp(-ik_1 a) = ir\sin k_1 a$ and $r_2 = -ir\sin k_2 b$. The product of the reflection coefficients is $r_1 r_2 = r^2 \sin k_1 a \sin k_2 b$. Taking this into account, we obtain

$$\cos K\Lambda = \cos kD - P/2, \quad (4)$$

where $P = 4r_1 r_2/(1 - r^2)$.

In the vicinity of the boundaries of forbidden gaps, we have $K\Lambda \approx m\pi$, where $m = 1, 2, \ldots$ is the number of the forbidden gap. We transform Eq. (4) as follows:

$$\cos(K\Lambda - m\pi) = \cos(kD - m\pi) - (-1)^m P/2$$
$$\approx 1 - (kD - m\pi)^2/2 - (-1)^m P/2$$
$$\approx \cos\sqrt{(kD - m\pi)^2 + (-1)^m P}.$$

In the allowed bands, we have $(kD - m\pi)^2 + (-1)^m P \geq 0$ and

$$(kD - m\pi)^2 \approx (K\Lambda - m\pi)^2 - (-1)^m P. \quad (5)$$

Since $K\Lambda = m\pi + ix$ and $(kD - m\pi)^2 + (-1)^m P < 0$ in forbidden gaps, we have

$$x \approx \sqrt{-(kD - m\pi)^2 - (-1)^m P}. \quad (6)$$

Dependences (5) and (6) correctly approximate Eq. (1) not only in the vicinity of the forbidden gap boundaries, but also in the entire range of K values. For example, for $r_Z = 2.7$ (which corresponds to $n_1 = 1.5$ and $n_2 = 4$ for PtCs) and $k_2 b/k_1 a = 3$ (alternation of layers with optical thickness $3\lambda_0/8$ and $\lambda_0/8$, where λ_0 is the wavelength at the mean frequency of the first forbidden

gap), the plots for dispersion characteristics calculated by formulas (1) and (5), (6) coincide.

Figure 1 shows the dispersion characteristics of a SL. The SL is formed by $Al_xGa_{1-x}As$ and GaAs layers with a thickness equal to 30 GaAs lattice constants in the [100] direction (equal to 2.82665 Å [12]); $V = 0.25$ eV; $m_1 = (0.0665 + 0.0835x)m_0$ [13]; $m_2 = 0.0665m_0$, where m_0 is the electron rest mass; V and x are connected via the relation $V = 0.7731x$ [14]. The approximation is also quite admissible at the beginning of the dependence, where $Z_1 \approx 0$ and $r \approx 1$. The maximal relative error in dependences 2 and 4 is less than 2%.

DISPERSION CHARACTERISTICS WITH EXPLICIT DEPENDENCE

The maximal width of forbidden gaps corresponds to quarter-wave layers $k_1a = k_2b$. In this case, $P = 4r^2\sin^2k_1a/(1 - r^2) \approx 4r^2$. As a result of simplification of relations (5) and (6), the dispersion characteristics for PtCs and PnCs become explicit. For allowed bands, we have

$$(kD - m\pi)^2 \approx (K\Lambda - m\pi)^2 + 4r^2, \qquad (7)$$

$$\omega \approx \frac{c}{D}[m\pi \mp \sqrt{(K\Lambda - m\pi)^2 + 4r^2}], \qquad (8)$$

where $m = 1, 3, \ldots$ (forbidden gaps with even m are absent); the minus sign corresponds to the first allowed band and high-frequency halves of the next allowed bands, while the plus sign corresponds to their low-frequency halves. The value of m in a high-frequency half of the allowed band increases by two as compared to its

Fig. 2. Dispersion characteristics of a CS. Curves *1, 2* and *3, 4* represent the real and imaginary parts of K in accordance with relations (1), (8) and (1), (9); $r_Z = 2.7$ (quarter-wave layers). Quantity ω is expressed in units of c/Λ; the real part of K is expressed in units of $1/\Lambda$ and the imaginary part, in arbitrary units.

value in the low-frequency half. In forbidden gaps, we have

$$\omega \approx \frac{c}{D}(m\pi \pm \sqrt{4r^2 - x^2}). \qquad (9)$$

The dispersion characteristics of a CS shown in Fig. 2 illustrate good agreement between Eqs. (8), (9) and (1).

It should be noted that dependence (7) corresponds to the dispersion characteristic of a waveguide, $k^2 = K^2 + K_m^2$, where K is the longitudinal wave number in the waveguide, K_m is the cutoff wave number, and m is the number of the waveguide mode. Such a correspondence is due to the universal nature of the physical phenomena determining the characteristics of periodic structures and waveguides, i.e., multibeam interference of reflected waves. The interference is constructive near forbidden gap boundaries and in the forbidden gaps themselves of the CS as well as in the vicinity of the cutoff frequencies of the waveguide. The latter frequencies correspond to the high-frequency boundaries of the forbidden gap. Common features of the spectral characteristics for waveguides and CSs may serve as the basis for application of model concepts and structural features of CSs similar to those used for waveguide structures.

IMPROVEMENT OF CS SELECTIVITY IN ODD OR EVEN ALLOWED BANDS

In the formation of PtC structures, use is made of the solutions developed for multilayer optical filters. One of such designs is based on periodic repetition of three-layer sections with alternating refractive indices of the layers, a quarter-wave inner layer, and outer layers with an optical thickness of $\lambda_0/8$ ($\lambda_0/8$ layer) [15]. As a result, a structure with quarter-wave inner layers and $\lambda_0/8$ outer layers is formed. Such a structure can also be used for other CSs.

Let us suppose that $r_Z > 1$ as before. A CS with $\lambda_0/8$ outer layers has an elevated selectivity in odd allowed bands as compared to a CS with quarter-wave outer layers if the impedance of the external medium is $Z_0 < Z_1$ or $Z_0 > Z_2$ (e.g., $Z_0 = Z_{1,2}$), while the impedance of the outer layers $Z_b = Z_{1,2}$, respectively, and in the even layers if $Z_0 > Z_1$ or $Z_0 < Z_2$, and $Z_b = Z_{1,2}$, respectively. Here, we take into account the combination of allowed bands due to the absence of forbidden gaps with even m. The proposed model makes it possible to visualize and optimize such an improvement of selectivity.

The inset to Fig. 3 shows a fragment of a CS with $\lambda_0/8$ outer layers. We divide the quarter-wave layer having impedance Z_1 and closest to the outer layer into halves by an imaginary boundary. We place antiphase mutually compensating sources of reflection with an amplitude equal to r on this boundary and sum the sources locates at the boundaries of the outer layer,

reducing them to the center $r_1 = i2r\sin(\pi F/4)$ of this layer, where $F = f/f_0$ is the normalized frequency and f_0 is the mean frequency of the first forbidden gap. Summing similar sources located at the boundaries of the left half of the quarter-wave layer with impedance Z_1 and reducing them to the axial line of this half, we also obtain a source with amplitude r_1. As a result of summation of two sources with amplitudes r_1 and their reduction to the middle of the layer with impedance Z_2, we obtain an equivalent reflector of period $r_\Lambda = 2r_1\cos(3\pi F/4)$. The sources of reflection located at the boundaries of the first half of a quarter-wave layer with impedance Z_1 and the left half of the next quarter-wave layer with impedance Z_1 also form an equivalent reflector of period with an amplitude equal to r_Λ. The frequency dependence of the equivalent reflector of period is defined by the expression $\hat{r}_\Lambda = |r_\Lambda|/4r = |\sin(\pi F/4)\cos(3\pi F/4)|$. The second cofactor ensures the main decrease in the sidelobe level of the frequency dependence of the low-frequency reflection coefficient for a CS in a range of variation of F from 0 to 2 with rejection at frequency $F_r = 2/3$ if the forbidden gap width $\Delta F \le F_r$. For $\Delta F > F_r$, the rejection point is displaced to the range of frequencies $F < F_r$. In a wider frequency range, the selectivity in odd allowed bands increases.

Figure 3 shows the frequency dependences of the reflection coefficient of a CS with $\lambda_0/8$ outer layers and an equivalent reflector of period. It can be seen that the decrease in the sidelobe level of dependence *1* is determined by dependence *2*. To reduce the ascent of the far (from the forbidden band) low-frequency sidelobes of dependence *1*, we must optimize the amplitudes of the reflectors at the CS boundaries by slightly increasing Z_0 for the first version and by reducing it for the second version. For $\hat{Z}_0 = Z_0/Z_2 = 1.11$ in the first version and $\hat{Z}_0 = 1.57$ in the second version, the near and far low-frequency sidelobes level out and the level of the far sidelobe decreases from 0.27 to 0.17.

In the case when $Z_0 > Z_1$ and $(Z_0 - Z_1)/(Z_0 + Z_1) = r$ (first version) or $Z_0 < Z_2$ and $(Z_2 - Z_0)/(Z_2 + Z_0) = r$ (second version), and $Z_b = Z_{1,2}$, respectively, the source of reflection located on the CS boundary coincides in amplitude and phase with the source of reflection located at the boundary between the extreme and adjacent layers. In this case, the amplitudes of the extreme equivalent CS reflector and the equivalent reflector of period are $r_1 = \mp 2\,r\cos(\pi F/4)$ and $r_\Lambda = \mp i\,2r_1\sin(3\pi F/4)$, respectively, where the minus and plus signs correspond to the above-mentioned versions. The frequency dependence of the equivalent reflector of period is defined by formula $\hat{r}_\Lambda = |\cos(\pi F/4)\sin(3\pi F/4)|$. Such a CS has an elevated selectivity at high frequencies in the range of F from 0 to 2 with rejection at frequency $F_r = $
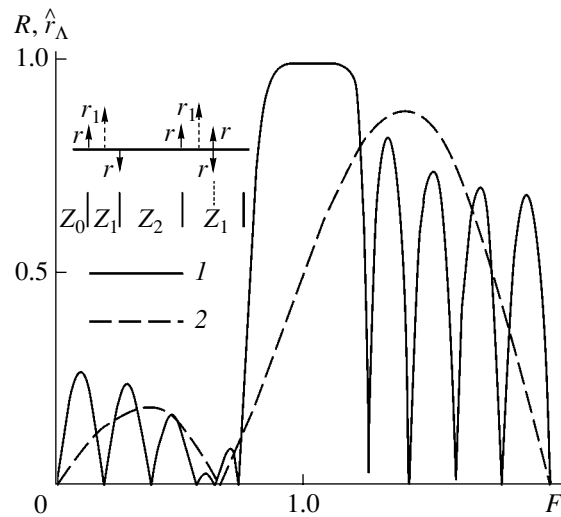

Fig. 3. Frequency dependences of the reflection coefficient R of a CS with $\lambda_0/8$ outer layers (*1*) and of an equivalent reflector of period $\hat{r}_\Lambda$ (*2*). The number of layers $N = 11$, $r_Z = 1.74$, $Z_0 = Z_2$, $Z_b = Z_1$ (first version) or $Z_0 = Z_1$, $Z_b = Z_2$ (second version).

4/3; in a wider range, the selectivity is high in even allowed bands.

ELEVATION OF CS SELECTIVITY IN ALL ALLOWED BANDS

The model of equivalent reflectors makes it possible to propose a basic solution ensuring elevated selectivity in all allowed bands.

Let us consider peculiarities of multibeam interference for the entire CS (see the inset to Fig. 1). We sum the sources located at the boundaries of quarter-wave layers for the CS period $r_\Lambda = i2r\sin(\pi F/2)$. Summing sources with r_1 and r_2, we obtain $r'_\Lambda = -2r\sin^2(\pi F/2)$. In allowed bands, interference is destructive and the pulsation level of the characteristics of CS reflection and transmission is determined by the frequency dependences of equivalent reflectors of period $\hat{r}_\Lambda = |r_\Lambda|/2r = |\sin(\pi F/2)|$ and $\hat{r}'_\Lambda = |r'_\Lambda|/2r = \sin^2(\pi F/2)$. In the second case, the pulsation level is noticeably lower. The realization of the second version of interference for the entire CS requires that the amplitudes of reflectors at its boundaries be halved by choosing the impedances $Z_{01,02}$ of the surrounding medium from the condition $Z_{01,02} = Z_{1,2}(Z_{1,2} + 3Z_{2,1})/(3Z_{1,2} + Z_{2,1})$, where $Z_{1,2}$ is the impedance of the layer bordering the medium.

Figure 4 illustrates the elevation of the efficiency of transmission of waves in allowed bands (as compared to the traditional solution) with the help of the compensation proposed above.

The model proposed here visually illustrates the key role of mutual compensation of reflections in the for-
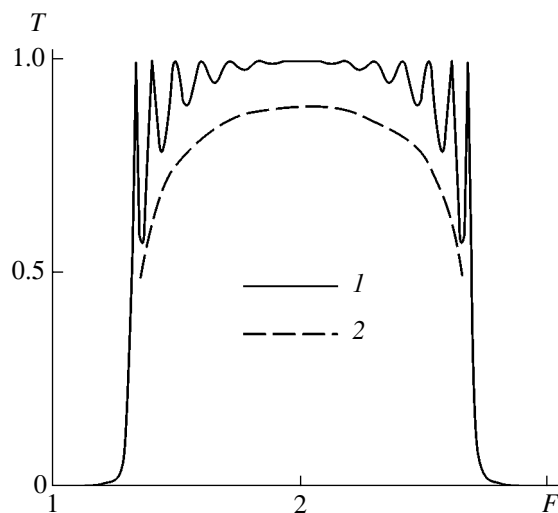
Fig. 4. Dependence of the transmission coefficient T of a CS: the solution proposed here (*1*) and traditional solution (*2*) (the envelope of pulsations is shown). $N = 15$, $r_Z = 2.7$, $\hat{Z}_0 = 1.6$, and $\hat{Z}_0 = 2.7$, respectively, for dependences *1* and *2*.

mation of the allowed band. The method of additional compensation of reflections based on this model makes it possible to noticeably improve the CS sensitivity, which is of utmost importance for high-efficiency CS-based devices.

REFERENCES

1. A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, New York, 1984; Mir, Moscow, 1987).

2. R. M. Ridder and R. Stoffer, AIP Conf. Proc. **560** (1), 99 (2001).

3. A. M. Kosevich and M. A. Mamalui, Zh. Éksp. Teor. Fiz. **122**, 897 (2002) [JETP **95**, 777 (2002)].

4. B. Gralak and D. Maystre, J. Mod. Opt. **47**, 1253 (2000).

5. S. M. Rytov, Zh. Éksp. Teor. Fiz. **29**, 605 (1955) [Sov. Phys. JETP **2**, 466 (1955)].

6. S. M. Rytov, Akust. Zh. **2**, 71 (1956) [Sov. Phys. Acoust. **2**, 68 (1956)].

7. A. M. Kosevich, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 633 (2001) [JETP Lett. **74**, 559 (2001)].

8. A. Yariv, *Introduction to Theory and Applications of Quantum Mechanics* (Wiley, New York, 1982; Mir, Moscow, 1984).

9. K. F. Brennan, IEEE Trans. Educ. **35**, 60 (1992).

10. A. N. Khondker, M. R. Khan, and A. F. M. Anwar, J. Appl. Phys. **63**, 5191 (1988).

11. A. F. M. Anwar, A. N. Khondker, and M. R. Khan, J. Appl. Phys. **65**, 2761 (1989).

12. T. K. Gaylord, E. N. Glytsis, and K. F. Brennan, J. Appl. Phys. **65**, 2535 (1989).

13. K. Nakamura, A. Shimizu, K. Fujii, *et al.*, IEEE J. Quantum Electron. **28**, 1670 (1992).

14. D. W. Wilson, E. N. Glytsis, and T. K. Gaylord, IEEE J. Quantum Electron. **29**, 1364 (1993).

15. A. J. Thelen, in *Physics of Thin Films: Advances in Research and Development,* Ed. by G. Haas, M. H. Francombe, and R. W. Hoffman (Academic, New York, 1974; Mir, Moscow, 1977), Vol. 5.

Translated by N. Wadhwa

SOLID-STATE
ELECTRONICS

Ramo–Shockley Relation for a Series *RCL* Circuit

N. A. Poklonski, S. A. Vyrko, and A. A. Kocherzhenko

Belarussian State University, Minsk, 220050 Belarus
e-mail: poklonski@bsu.by
Received March 3, 2004

Abstract—The current induced by the passage of an external point charge through a plane vacuum capacitor in an *RCL* circuit free of current (voltage) sources is calculated. The case is also analyzed when an internal point charge is emitted by one of the capacitor plates, moves to the other plate, and is absorbed by it. A technique is proposed to measure the internal charge and its velocity component perpendicular to the capacitor plates in a passive *RCL* circuit. © 2004 MAIK "Nauka/Interperiodica".

INTRODUCTION

The Ramo–Shockley relation states that external point charge Q moving perpendicularly to the plates of a vacuum capacitor with speed v induces a rectangular pulse of current $I_{RS} = Qv/b$ (where b is the distance between the capacitor plates) in a zero-inductance and zero-resistance circuit that closes the capacitor [1–3]. Current I_{RS} exists in the circuit only during the time $0 \leq t \leq b/v$, when the charge moves between the capacitor plates; $t_c = b/v$ is the transit time of charge Q. The Ramo–Shockley relation was generalized directly from the Maxwellian electrodynamics to the case in which the quasi-electrostatic approximation is invalid [4].

In [5], the Ramo–Shockley relation was generalized to a series *RC* circuit. However, in addition to capacitance and resistance, a real circuit has an inductance. Also, it is necessary to distinguish between the passage of an external charge through the capacitor and the transition of an internal charge from one plate of the capacitor to the other, for example, as a result of exoelectron emission [6, 7] (arrival of a charge to the second plate of the capacitor can be recorded from the reverse photoemission [8, 9]). The passage of an external charge through the capacitor in *RC* and *RL* circuits connected in parallel was considered in [10]. However, the input resistance of the circuit was disregarded in the final expressions.

The purpose of this study is to generalize the Ramo–Shockley relation to the series *RCL* circuit containing no current (voltage) sources for two cases: (i) external charge Q moves through the plane capacitor and (ii) internal charge Q is emitted by one plate of the uncharged capacitor, moves to the other plate, and is absorbed by it.

Consider a circuit consisting of capacitor *C*, resistor *R*, and inductor *L* connected in series (Fig. 1). The velocity v of charge Q between the capacitor plates is assumed to be constant. The lag of the induced electromagnetic field and its effect on the charge motion are neglected (see, e.g., [11]). The time origin ($t = 0$) is chosen to be the moment when the charge flies through the inner surface of the first capacitor plate it crosses. The current and voltage are considered for $0 \leq t \leq t_c$, when the external or internal charge moves between the plates, and for $t > t_c$, when the current is the discharge current in the *RCL* circuit.

EXTERNAL CHARGE

As the external charge moves ($0 \leq t \leq t_c$) between the plates of capacitor *C* of a series *RCL* circuit (Fig. 1), the following balance of currents holds:

$$I = I_1 = \frac{Qv}{b} + I_d = \frac{Qv}{b} - C\frac{dV}{dt}, \qquad (1)$$

where $I_1 = I$ is the current in the circuit, $Qv/b = I_{RS}$ is the plate-to-plate charge transport current, v is the charge's velocity component perpendicular to the plates, V is the electric potential difference between the plates induced by the moving external charge, and $I_d = -CdV/dt$ is the displacement current [12].

Potential difference V across the capacitor due to current I_1 excited in the *RCL* circuit is a sum of the
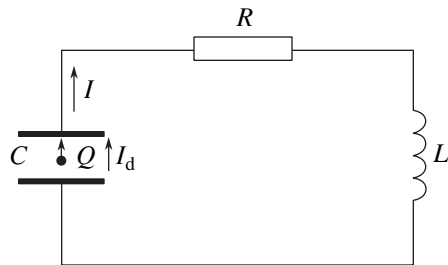

Fig. 1. Motion of a point charge *Q* between the plates of a vacuum plane capacitor in a passive *RCL* circuit: *I* is the current in the *RL* circuit and I_d is the displacement current in the capacitor.

potential differences across the resistor and inductor:

$$V = I_1 R + L \frac{dI_1}{dt}. \tag{2}$$

Substituting (2) into (1), we obtain the equation for the current (refer, e.g., to [13, 14]):

$$LC \frac{d^2 I_1}{dt^2} + CR \frac{dI_1}{dt} + I_1 = \frac{Qv}{b}; \tag{?}$$

which, in terms of the designations

$$\omega_0 = \frac{1}{\sqrt{LC}}, \quad \xi = \frac{R}{2} \sqrt{\frac{C}{L}}, \quad I_{RS} = \frac{Qv}{b},$$

becomes

$$\frac{d^2 I_1}{dt^2} + 2\xi\omega_0 \frac{dI_1}{dt} + \omega_0^2 I_1 = \omega_0^2 I_{RS}. \tag{3}$$

The general solution to Eq. (3) is a sum of the general solution I_0 to this equation with the zero right-hand part and any particular solution to this equation. As the particular solution to (3), we can take the constant current $I_{RS} = Qv/b$.

When $\xi \neq 1$, the solution to the homogeneous equation corresponding to (3) has the form

$$\begin{aligned} I_0 &= A_1 \exp(-\xi\omega_0 t + \omega_0 t \sqrt{\xi^2 - 1}) \\ &+ A_2 \exp(-\xi\omega_0 t - \omega_0 t \sqrt{\xi^2 - 1}), \end{aligned} \tag{4}$$

where A_1 and A_2 are the integration constants.

Then, the general solution to Eq. (3) is

$$\begin{aligned} I_1 &= I_0 + I_{RS} = A_1 \exp(-\xi\omega_0 t + \omega_0 t \sqrt{\xi^2 - 1}) \\ &+ A_2 \exp(-\xi\omega_0 t - \omega_0 t \sqrt{\xi^2 - 1}) + I_{RS}, \end{aligned} \tag{5}$$

where constants A_1 and A_2 are determined by the initial conditions

$$I_1|_{t=0} = 0, \quad \left.\frac{dI_1}{dt}\right|_{t=0} = 0, \tag{6}$$

which mean that, at the initial instant ($t = 0$), the electric current is zero and $V|_{t=0} = 0$ (see formula (2)).

Equality (5) with (6) takes the form

$$\begin{aligned} I_1 &= I_{RS} \Biggl\{ 1 - \Biggl[\cosh(\omega_0 t \sqrt{\xi^2 - 1}) \\ &+ \frac{\xi}{\sqrt{\xi^2 - 1}} \sinh(\omega_0 t \sqrt{\xi^2 - 1}) \Biggr] \exp(-\xi\omega_0 t) \Biggr\}. \end{aligned} \tag{7}$$

To obtain the result at $\xi = 1$, we pass to the limit $\xi \longrightarrow 1$ in (7) to get

$$I_1 = I_0 + I_{RS} = I_{RS}[1 - (1 + \omega_0 t)\exp(-\omega_0 t)]. \tag{8}$$

When $L \longrightarrow 0$, the circuit's eigenfrequency ω_0 and the parameter ξ tend to infinity with $\xi/\omega_0 \longrightarrow RC/2$. Then, formula (7) reduces to the known relation [5]

$$I_1 = I_{RS}\Biggl[1 - \exp\Bigl(-\frac{t}{RC}\Bigr) \Biggr], \tag{9}$$

where $0 \leq t \leq t_c$.

After the external charge passes through the capacitor (for $t > b/v = t_c$), current $I_1 = I$ in the circuit is described by Eq. (3) with $Q = 0$:

$$\frac{d^2 I_1}{dt^2} + 2\xi\omega_0 \frac{dI_1}{dt} + \omega_0^2 I_1 = 0 \tag{10}$$

under the conditions of current continuity (at $t = t_c$):

$$\begin{aligned} I_1|_{t_c - 0} &= I_1|_{t_c + 0}, \quad V|_{t_c - 0} = V|_{t_c + 0} \quad \text{or} \\ (dI_1/dt)_{t_c - 0} &= (dI_1/dt)_{t_c + 0}. \end{aligned} \tag{11}$$

At $\xi = 1$, the solution to Eq. (10) under conditions (11) has the simplest form

$$I_1 = (B_1 + B_2 t)\exp(-\omega_0 t), \tag{12}$$

where constant $B_1 = I_{RS}(1 - \omega_0 t_c)\exp(\omega_0 t_c)$ and $B_2 = I_{RS}[\exp(\omega_0 t_c) - 1]$.

The time dependence of current $I = I_1$ at $\xi = 1$ when the external charge Q moves through the capacitor and after that is shown in Fig. 2a.

INTERNAL CHARGE

When internal charge Q emitted by one plate of the uncharged capacitor in the *RCL* circuit moves to the other plate ($0 \leq t \leq t_c$) and is absorbed by the latter, the current in the circuit can be represented as the sum $I = I_1 + I_2$, where I_1 is the current induced when the charge moves between the plates and I_2 is the capacitor discharge current. As in the case of the external charge moving through the capacitor, the current I_1 is given by formulas (7) and (8) for $\xi \neq 1$ and $\xi = 1$, respectively. After emitting the charge Q, the capacitor becomes charged and starts discharging the discharge current I_2, which satisfies the equation (compare to (10))

$$\frac{d^2 I_2}{dt^2} + 2\xi\omega_0 \frac{dI_2}{dt} + \omega_0^2 I_2 = 0, \tag{13}$$

where $2\xi = RC\omega_0$.

The solution to Eq. (13) at $I_2|_{t=0} = 0$ and $L(dI_2/dt)|_{t=0} = Q/C$ is given by the formula (refer, e.g.,

to [13])

$$I_2 = \frac{Q\omega_0}{\sqrt{\xi^2 - 1}}\sinh(\omega_0 t\sqrt{\xi^2 - 1})\exp(-\xi\omega_0 t). \quad (14)$$

The total current in the circuit is the sum of (7) and (14):

$$I = I_1 + I_2 = I_{RS} - I_{RS}\left[\cosh(\omega_0 t\sqrt{\xi^2 - 1})\right.$$

$$\left. + \frac{\xi - \omega_0 t_c}{\sqrt{\xi^2 - 1}}\sinh(\omega_0 t\sqrt{\xi^2 - 1})\right]\exp(-\xi\omega_0 t). \quad (15)$$

At $\xi = 1$, the discharge current obtained from (14) is

$$I_2 = Q\omega_0^2 t\exp(-\omega_0 t), \quad (16)$$

and the total current is the sum of (8) and (16):

$$I = I_1 + I_2$$
$$= I_{RS} - I_{RS}[1 + (1 - \omega_0 t_c)\omega_0 t]\exp(-\omega_0 t). \quad (17)$$

When $L \longrightarrow 0$, i.e., for a series RC circuit, formula (15) yields for $0 \le t \le b/v = t_c$:

$$I = \frac{I_{RS}}{RC}\left[(RC + t_c) - (RC - t_c)\exp\left(-\frac{t_c}{RC}\right)\right]$$
$$\times \exp\left(\frac{t_c - t}{RC}\right). \quad (18)$$

Comparison of formulas (18) and (9) shows that the current in the RC circuit induced by the passage of the internal charge from one capacitor plate to another, other conditions being the same, is higher than the current induced by the passage of the external charge through the capacitor.

After internal charge Q emitted by one capacitor plate passes through the capacitor ($t > b/v = t_c$) and is absorbed by the other plate, the current in the circuit satisfies Eq. (10) under the conditions for the time $t = t_c$ (compare to (11)),

$$I|_{t_c - 0} = I|_{t_c + 0}, \quad V|_{t_c - 0} + Q/C = V|_{t_c + 0},$$

or, taking into account (2),

$$I|_{t_c - 0} = I|_{t_c + 0}, \quad (dI/dt)_{t_c - 0} + Q\omega_0^2 = (dI/dt)_{t_c + 0}, (19)$$

i.e., the current is a continuous but nonsmooth, function of time.

Solutions to (10) for $t > t_c$ have the form of (4) and (12), where, however, the coefficients A_1, A_2, B_1, and B_2 are found from conditions (19). The simplest case is that of $\xi = 1$. At $R^2C = 4L$ and $t > t_c$, the current induced in the RCL circuit by the passage of charge Q emitted by one capacitor plate and absorbed by the other plate
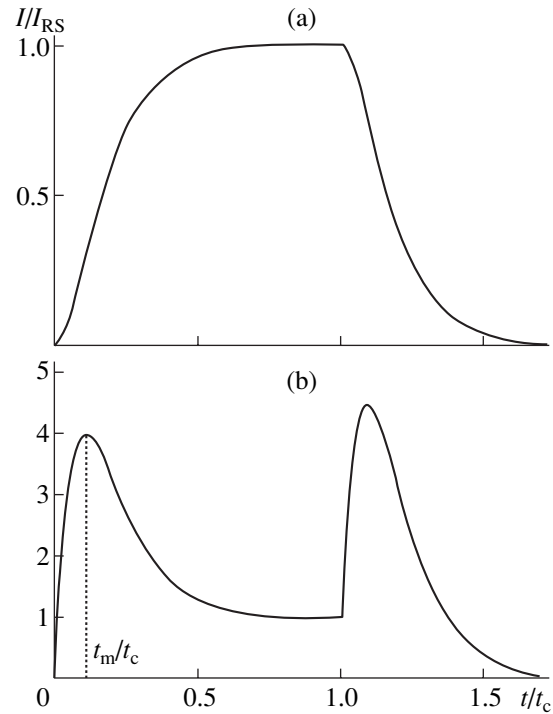
Fig. 2. Current I (normalized by $I_{RS} = Qv/b$) in the RCL circuit vs. time (normalized by $t_c = b/v$) at $RC = 4L/R$ ($\xi = 1$) for (a) the external charge Q passing through the capacitor with the velocity component perpendicular to its plates, v, and (b) the charge Q with the same velocity component v emitted by one capacitor plate and absorbed by the other plate after the time t_c; $t_m/t_c = (t_c/\sqrt{LC} - 1)^{-1}$.

is described by formula (12), where the coefficients B_1 and B_2 are calculated with (19) as

$$B_2 = \omega_0\{[I_{RS} + Q\omega_0]\exp(\omega_0 t_c)$$
$$- I_{RS}[1 + 3\omega_0 t_c - 2\omega_0^2 t_c^2]\},$$

$$B_1 = I_{RS}\exp(\omega_0 t_c) - I_{RS}[1 + \omega_0 t_c - \omega_0^2 t_c^2] - B_2 t_c.$$

The time dependence of the current $I = I_1 + I_2$ in the RCL circuit during the passage of the internal charge between the capacitor plates and after that at $\xi = 1$ is shown in Fig. 2b.

DISCUSSION

Comparison of formulas (7) with (15) and (8) with (17) shows that the external charge Q, moving through the capacitor, and the internal charge Q, emitted by one capacitor plate and absorbed by the other plate, induce different currents in the RCL circuit. This is because, when the internal charge is emitted, the capacitor becomes charged, which produces the discharge current I_2. The lower the capacitance C of the capacitor, the higher this current. The total current in the series RCL circuit at the initial instant ($t = 0$) is zero, because

the circuit contains an inertial element—the inductor L. Note that, in a series RC circuit for the case of internal charge Q, the current abruptly changes from 0 to $I = Q/RC$ at the time when the charge is emitted (at $t = 0$).

Let us show that taking into account the inertial properties of the RCL circuit makes it possible to measure the velocity component v of the internal charge Q (emitted by one plate and moving toward the other plate) perpendicular to the capacitor plates. The simplest situation is at $\xi = 1$, when the time dependence of the current I in the RCL circuit is described by formula (17) and the potential difference (voltage) across the inductor L is given by the expression

$$V_L = L\frac{dI}{dt} = LI_{RS}\omega_0^2[(1-\omega_0 t_c)t + t_c]\exp(-\omega_0 t).\text{(20)}$$

According to (20), the voltage V_L vanishes at the time $t_m = t_c/(t_c\omega_0 - 1)$ (at $t = t_m$, the current I in the RCL circuit reaches its maximum value I_m). If $0 < t_m < b/v = t_c$ (i.e., the transit time $t_c = b/v$ of charge Q between the capacitor plates is long enough for the current in the RCL circuit to reach its maximum value I_m), the charge's velocity, according to (20), is

$$v = b\left(\frac{1}{t_m} - \omega_0\right) = b\left(\frac{1}{t_m} - \frac{1}{\sqrt{LC}}\right), \qquad (21)$$

where the time t_m is measured experimentally from the condition $V_L(t_m) = 0$.

If the total current I in the circuit (or the voltage $U_R = IR$ across the resistor R) is measured simultaneously, the internal charge moving between the capacitor plates can also be found from (17):

$$Q = \frac{I_m t_m}{1 + \omega_0 t_m(1 - 2\omega_0 t_m)[\exp(-\omega_0 t_m) - 1]}, \quad (22)$$

where $I_m = U_R(t_m)/R$.

Note that the time t_m, when the voltage across the inductor vanishes, $V_L(t_m) = 0$, depends only on the velocity component v of the internal charge Q perpendicular to the capacitor plates and is independent of the charge.

Boundary conditions (19) assume that the charge emitted by one plate is absorbed by the other plate, after which the current in the circuit increases again (due to the capacitor discharge current I_2) and maximizes once more. To calculate v and Q from formulas (21) and (22), one should use the first maximum of the current (formulas using the second maximum are more complex).

Formulas that give the charge and its velocity at $\xi > 1$ can be derived in a similar way; however, these formulas are more complex than (21) and (22). When $\xi < 1$, it is difficult to measure v and Q, because the current in this case is oscillatory and can have several extrema even in the time interval from 0 to t_c.

The condition $\xi = 1$ ($RC = 4L/R$) is equivalent to the situation in which the resistance of a circuit consisting of parallel-connected RC and RL circuits, that have resistors with equal resistances $R/2$, is purely active and equals $R/2$ at any frequency (see, e.g., [15, 16]). This circumstance can be used to choose parameters of the series RCL circuit.

Note that formula (21) is only valid for an internal charge (emitted by one of the capacitor plates). For an external charge, the current I_1 in the RCL circuit is described by formula (8) and has no extrema for $0 \le t \le b/v$.

Thus, expressions for the current I induced in the RCL circuit containing no current (voltage) sources by the external charge Q passing through the capacitor or by an internal charge Q, emitted by one capacitor plate and absorbed by the other plate, are obtained. A method for measuring the charge-velocity component v perpendicular to the capacitor plates and the internal charge Q in the source-free circuit is proposed.

REFERENCES

1. W. Shockley, J. Appl. Phys. **9**, 635 (1938).
2. S. Ramo, Proc. IRE **27**, 584 (1939).
3. P. De Visschere, Solid-State Electron. **33**, 455 (1990).
4. P. D. Yoder, K. Gärtner, and W. Fichtner, J. Appl. Phys. **79**, 1951 (1996).
5. A. N. Inovenkov, O. V. Konstantinov, and V. I. Pirogov, Zh. Tekh. Fiz. **63** (9), 1 (1993) [Tech. Phys. **38**, 739 (1993)].
6. E. Rabinovich, Usp. Fiz. Nauk **127**, 163 (1979).
7. G. A. Mesyats, *Ectons in a Vacuum Discharge: Breakdown, the Spark, and the Arc* (Nauka, Moscow, 2000) [in Russian].
8. O. M. Artamonov and S. N. Samarin, Zh. Tekh. Fiz. **61** (10), 186 (1991) [Sov. Phys. Tech. Phys. **36**, 1186 (1991)].
9. P. D. Johnson and S. L. Hulbert, Rev. Sci. Instrum. **61**, 2277 (1990).
10. N. A. Poklonski, V. V. Mityanok, and S. A. Vyrko, Pis'ma Zh. Tekh. Fiz. **28** (15), 33 (2002) [Tech. Phys. Lett. **28**, 635 (2002)].
11. M. A. Miller, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **29**, 991 (1986).
12. E. M. Purcell, *Electricity and Magnetism*, 2nd ed. (McGraw-Hill, New York, 1985; Nauka, Moscow, 1975), p. 272.
13. T. A. Tatur and V. E. Tatur, *Steady-State and Transient Processes in Electrical Circuits* (Vysshaya Shkola, Moscow, 2001) [in Russian].
14. G. Magetto, *Le thyristor: definitions, protections, commandes* (Presses universitaires de Bruxelles, Bruxelles, 1971; Énergiya, Moscow, 1977).
15. L. A. Vaĭnshteĭn, Usp. Fiz. Nauk **118**, 339 (1976) [Sov. Phys. Usp. **19**, 189 (1976)].
16. Yu. S. Barash and V. L. Ginzburg, Usp. Fiz. Nauk **118**, 523 (1976) [Sov. Phys. Usp. **19**, 263 (1976)].

Translated by A. Khzmalyan

OPTICS,
QUANTUM ELECTRONICS

Two-Wavelength Dispersion Holographic Lateral-Shift Interferometry

A. M. Lyalikov

Grodno State University, Grodno, 230023 Belarus
e-mail: lyalikov@inbox.ru
Received July 16, 2003; in final form, April 26, 2004

Abstract—It is demonstrated that two-wavelength holographic interferometry with a small lateral shift in a grating interferometer makes it possible to study the dispersion characteristics of transparent objects using probe beams with arbitrary wavelengths. Interference patterns reconstructed represent fringes on the reconstructed image of the object, which characterize the value of the derivative of the difference between the refractive indices of the medium under study at the probe wavelengths along the direction of the shift. The results of experiments employing the method proposed are presented. © 2004 MAIK "Nauka/Interperiodica".

INTRODUCTION

Dispersion two-wavelength holographic interferometry is widely used for the purposes of plasma diagnostics, where it makes possible determining the electron concentration in the absence of corrections related to the refraction of heavy particles [1–7]. Detection of two-wavelength holograms under nonlinear conditions also makes possible the reconstruction of interferograms in cases when the interference pattern depends on the difference $n_1(x, y, z) - n_2(x, y, z)$ between the refractive indices at wavelengths λ_1 and λ_2 of the probe beams [1, 2]. The measurements of interference patterns corresponding to the lines of constant difference between the refractive indices using the method proposed was realized for integer ratios of wavelengths (in particular, for $\lambda_2/\lambda_1 = 2$). This was related to the dependence of the coefficient of the sensitivity of measurements on the diffraction order in the nonlinear hologram. In contrast to two-beam interferometry based on the comparison of a probe wave front with the front of a reference wave, the method of interferometry using small lateral shift makes it possible to measure interferograms with the sensitivity of measurements depending on the relative shift Δs of the interfering wave fronts [8]. Lateral-shift interferometry is a promising diagnostic tool for the systems in which it is difficult to form a reference wave.

Recently, shift interferometry has been widely used in the measurements of small distortions of wave fronts [9, 10]. The application of holographic principles makes it possible to significantly increase the efficiency of lateral-shift interferometry [7, 11–14]. Based on various methods for detecting and optical processing of holographic lateral-shift interferograms, which are widely used in holographic interferometry of phase objects, it is possible to develop methods for studying fast processes in transparent objects with a wide varia-

tion in the coefficient of the sensitivity of measurements [15, 16].

This paper demonstrates the possibility of studying the dispersion characteristics of transparent media at arbitrary wavelengths by means of interferometry with small lateral shift using a grating interferometer.

RECORDING TWO-WAVELENGTH HOLOGRAPHIC LATERAL-SHIFT INTERFEROGRAMS

Figure 1a shows the optical scheme of a two-wavelength setup for studying the dispersion characteristics of transparent media. The principal component of the setup is a grating interferometer [17] capable of recording holographic lateral-shift interferograms.

Light beams generated by lasers *1* and *2* are coaligned using beamsplitter *3* and directed to the shift interferometer equipped with a grating. Two beams expanded and collimated using a telescope consisting of lenses *4* and *5* are incident on object *6* under study. The variations $\varphi_1(x, y)$ and $\varphi_2(x, y)$ in the phases of light beams passing through transparent object *6* can be represented as

$$\varphi_{1,2}(x, y) = \frac{2\pi}{\lambda_{1,2}} \int_0^l n_{1,2}(x, y, z)\,dz, \qquad (1)$$

where subscripts *1* and *2* correspond to the first and second beams with wavelengths λ_1 and λ_2, respectively; $n_1(x, y, z)$ and $n_2(x, y, z)$ are the refractive indices of object *6* at wavelengths λ_1 and λ_2, respectively; and *l* is the thickness of object *6* along the direction of probing. Cartesian coordinates *xyz* are chosen in such a way that the probe beam is parallel to the *z* axis.
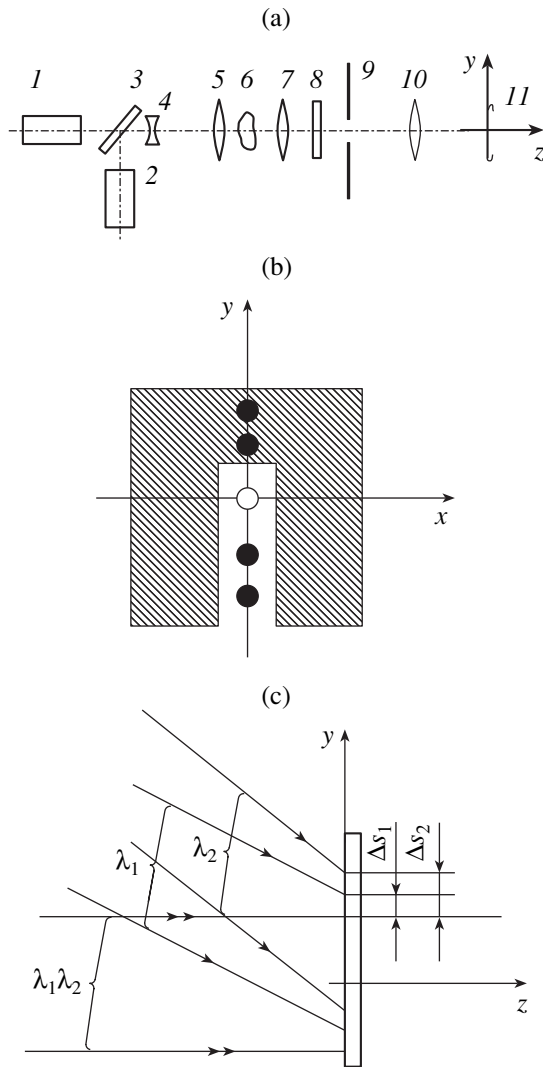
(a)

(b)

(c)

Fig. 1. (a) Optical scheme of the setup for studying the dispersion characteristics of transparent media, (b) configuration of aperture and diffraction spectrum of light beams with wavelengths λ_1 and λ_2, and (c) the scheme of beam overlapping in the detection plane of two-wavelength holographic lateral-shift interferogram.

Each light beam is split into two coherent beams using grating *8* placed in the vicinity of the back focal plane of objective *7* in front of aperture *9*. We assume that the grooves of grating *8* are perpendicular to the *y* axis. Figure 1 shows aperture *9* and the diffraction spectrum of light beams with wavelengths λ_1 and λ_2. Diffraction maxima are situated on a straight line parallel to the *y* axis. The central maximum corresponds to the zero order diffraction of beams with wavelengths λ_1 and λ_2. Note that the zero-order diffraction maxima of two beams coincide, while the diffraction maxima of +1 and –1 orders are spatially separated so that the greater the wavelength, the greater the displacement along the *y* axis (we assume that $\lambda_1 < \lambda_2$). The configuration of the aperture (Fig. 1b) makes it possible to sep-

arate the beams diffracted to the order of a certain sign (e.g., –1 order). Thus, two pairs of beams with wavelengths λ_1 and λ_2 are separated from other beams by aperture *9* and collimated with objective *10*. In plane *11*, we observe two lateral-shift interference patterns corresponding to the two wavelengths. Figu-re 1c illustrates the superposition of light beams at the detection plane *11* giving rise to two-wavelength holographic lateral-shift interferogram. The lateral shifts Δs_1 and Δs_2 of the interference patterns formed by the light beams with wavelengths λ_1 and λ_2 are directed along the *y* axis and differ from each other owing to the dispersion of grating *8*. At small shifts satisfying condition $\Delta s_{1,2} \ll L$ (*L* is the transverse size of object *6* under study), the intensity distribution in the holographic lateral-shift interferograms is represented as

$$I_{1,2}(x, y) \sim 1 + \cos\left[\frac{2\pi y}{T} + \Delta s_{1,2}\frac{\partial \varphi_{1,2}(x, y)}{\partial y}\right], \quad (2)$$

where *T* is the period of fringes in holographic interferograms depending on the angles between the interfering coherent beams. This period can be controlled by changing the period of diffraction grating *8* and its position relative to the back focal plane of objective *7* [18, 19].

Note that the periods of fringes in the holographic interferograms obtained using beams with different wavelengths are equal, owing to the features of the lateral-shift interferometer with diffraction grating [19].

Thus, a pair of holographic lateral-shift interferograms (2) is detected at plane *11*. Owing to the difference between wavelengths λ_1 and λ_2, these interferograms are incoherently superimposed on a film, which gives rise to a two-wavelength holographic lateral-shift interferogram. When the film is exposed under linear conditions (the contrast ratio of photoemulsion is $\gamma = -2$) and chemically processed, the amplitude transmittance of the two-wavelength holographic lateral-shift interferogram is represented as

$$\tau(x, y) \sim I_1(x, y) + I_2(x, y) \sim 2$$

$$+ \cos\left[\frac{2\pi y}{T} + \Phi_1(x, y)\right] + \cos\left[\frac{2\pi y}{T} + \Phi_2(x, y)\right], \quad (3)$$

where

$$\Phi_{1,2}(x, y) = \Delta s_{1,2}\frac{\partial \varphi_{1,2}(x, y)}{\partial y}. \quad (4)$$

The amplitude transmittance of a two-wavelength holographic interferogram given by expression (3) is similar to the transmittance of a double-exposure hologram.

RECONSTRUCTION OF CONTOUR LINES

A method of reconstruction of the interference pattern from the two-wavelength holographic lateral-shift

interferogram (3) is basically the same as the method used for reconstructing interference patterns from double-exposure holograms. When hologram (3) is illuminated with a collimated light beam, two waves propagate in the first diffraction order. These waves are diffracted by the fringes of holographic structures corresponding to the holographic lateral-shift interferograms formed by the beams with wavelengths λ_1 and λ_2. In accordance with expression (3), the phases of these waves are $\Phi_1(x, y)$ and $\Phi_2(x, y)$. A conventional method to increase the visibility of the reconstructed contour lines involves filtering of the spatial frequencies (an opening in the aperture separates light beams diffracted in the first order). In this case, it follows from expression (3) that the intensity distribution in the interference pattern can be written as

$$I_2(x, y) \sim 1 + \cos[\Phi_1(x, y) - \Phi_2(x, y)]. \qquad (5)$$

For light interference fringes, the following condition is satisfied:

$$\Phi_1(x, y) - \Phi_2(x, y) = 2\pi N, \text{ where } N = 0, 1, 2, \dots. \quad (6)$$

Taking into account expressions (1) and (4), we can represent expression (6) as

$$\frac{\Delta s_1}{\lambda_1}\frac{\partial}{\partial y}\left[\int_0^l n_1(x, y, z)dz\right] - \frac{\Delta s_2}{\lambda_2}\frac{\partial}{\partial y}\left[\int_0^l n_2(x, y, z)dz\right] = N. (7)$$

This expression becomes even more convenient when the coefficients characterizing the sensitivity of measurements are equal:

$$\frac{\Delta s_1}{\lambda_1} = \frac{\Delta s_2}{\lambda_2} = C. \qquad (8)$$

One of the remarkable features of the lateral-shift interferometer with diffraction grating placed in the focal plane or in the vicinity of focus consists in that the ratio of the shift of interfering beams to the wavelength of light source remains constant: $\Delta s / \lambda = \text{const}$ [19].

Thus, condition (8) is always satisfied for the case under consideration. With allowance for this condition, we can represent the condition for the formation of light interference fringes as

$$\frac{\partial}{\partial y}\left\{\int_0^l [n_1(x, y, z) - n_2(x, y, z)]dz\right\} = \frac{N}{C}. \qquad (9)$$

Expression (9) describes the interference fringes in the reconstructed image of the object under study. Each interference fringe corresponds to a zone of the transparent object under study where the product of the derivative of the difference between the refractive indices and the thickness of the object is constant. The interference pattern depends on the dispersion characteristics of the object under study.
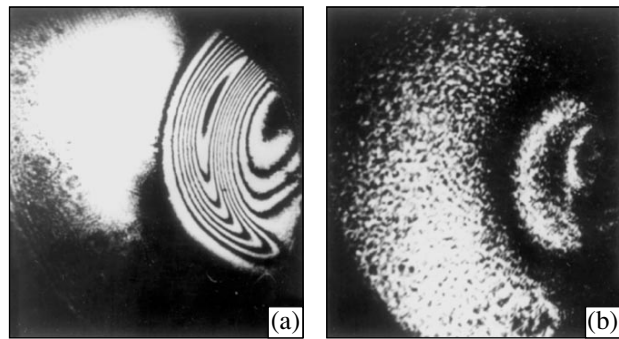
Fig. 2. (a) Lateral-shift interferogram ($\Delta s = 1.0$ mm) of a glass plate obtained using the radiation of helium–neon laser ($\lambda_2 = 632.8$ nm) upon tuning to an infinitely wide fringe and (b) interference pattern reconstructed from the two-wavelength holographic lateral-shift interferogram of the same glass plate recorded using the probe beams with wavelengths of $\lambda_1 = 488.0$ nm and $\lambda_2 = 632.8$ nm.

EXPERIMENTAL TESTS

In order to verify the proposed method, we used an experimental setup corresponding to the scheme depicted in Fig. 1a. Lasers *1* and *2* were an argon laser with the wavelength $\lambda_1 = 488.0$ nm and an LGN-215 helium–neon laser with the wavelength $\lambda_2 = 632.8$ nm, respectively. To equalize the exposures at two wavelengths in the course of recording of two-wavelength holographic lateral-shift interferogram, we rotated a polarizer at the exit of the argon laser. Grating *8* (Fig. 1a) was a phase diffraction grating with a period of about 60 μm recorded on an FG-690 film using a holographic method with subsequent bleaching.

The object under study was a glass plate. The surface of this plate was not flat as a result of thermal processing. Figure 2a shows the lateral-shift interferogram ($\Delta s = 1.0$ mm) of the plate under study obtained using the radiation of the helium–neon laser ($\lambda_2 = 632.8$ nm) upon tuning to an infinitely wide fringe. Figure 2b demonstrates the interference pattern reconstructed from the two-wavelength holographic lateral-shift interferogram of the same glass plate.

When the two-wavelength holographic lateral-shift interferogram was recorded on FG-690 film, the relative shifts of the beams with wavelengths of $\lambda_1 = 488.0$ nm and $\lambda_2 = 632.8$ nm were $\Delta s_1 = 0.8$ mm and $\Delta s_2 = 1.0$ mm, respectively. In the interferogram shown in Fig. 2b, the interference fringes correspond to the zones of the glass plate under study where the product of the derivative of the difference between the refractive indices at wavelengths of 488.0 and 632.8 nm along the direction of the shift and the thickness of glass is constant. The interferogram characterizes the surface relief of the glass plate, and its sensitivity depends on the dispersion characteristics of glass.

CONCLUSIONS

The proposed method will be effective in the study of the dispersion characteristics of plasma objects. Any pulsed light source with wavelengths suitable for plasma probing can serve as a source of the probe beam.

REFERENCES

1. G. V. Ostrovskaya and Yu. I. Ostrovskiǐ, Zh. Tekh. Fiz. **40**, 2419 (1970) [Sov. Phys. Tech. Phys. **15**, 1890 (1970)].

2. A. B. Ignatov, I. I. Komissarova, G. V. Ostrovskaya, *et al.*, Zh. Tekh. Fiz. **41** (1971) [Sov. Phys. Tech. Phys. **16** (1971)].

3. G. V. Ostrovskaya, in *Proceedings of the 4th All-Union School on Holography, Leningrad, 1973*, pp. 51–76.

4. G. V. Ostrovskaya and N. P. Pobedonostseva, Zh. Tekh. Fiz. **45**, 1462 (1975) [Sov. Phys. Tech. Phys. **20**, 923 (1975)].

5. Kh. P. Alum, Yu. V. Koval'chuk, and G. V. Ostrovskaya, in *Proceedings of the 13th All-Union School on Holography, Leningrad, 1981*, pp. 84–97.

6. Kh. P. Alum, Yu. V. Koval'chuk, and G. V. Ostrovskaya, Zh. Tekh. Fiz. **54**, 896 (1984) [Sov. Phys. Tech. Phys. **29**, 534 (1984)].

7. C. M. Vest, *Holographic Interferometry* (Wiley, New York, 1979; Mir, Moscow, 1982).

8. *Optical Shop Testing*, Ed. by D. Malacara (Wiley, New York, 1978; Mashinostroenie, Moscow, 1985).

9. A. M. Nugumanov, R. V. Smirnov, and V. I. Sokolov, Kvantovaya Élektron. (Moscow) **30**, 435 (2000).

10. V. I. Sokolov, Kvantovaya Élektron. (Moscow) **31**, 891 (2001).

11. V. G. Kulkarni, Opt. Laser Technol. **5**, 269 (1979).

12. V. G. Gusev, Opt. Mekh. Prom., No. 4, 3 (1992).

13. A. F. Tuev, Opt. Mekh. Prom., No. 3, 75 (1993).

14. V. G. Gusev, Izv. Vyssh. Uchebn. Zaved. Fiz. **43** (12), 8 (2000).

15. A. M. Lyalikov, Opt. Spektrosk. **93**, 512 (2002) [Opt. Spectrosc. **93**, 472 (2002)].

16. A. M. Lyalikov, Opt. Spektrosk. **93**, 517 (2002) [Opt. Spectrosc. **93**, 477 (2002)].

17. A. F. Belozerov and N. M. Spornik, Tr. Voenno-Vozdushn. Inzh. Akad. im. Zhukovskogo, No. 1301, 90 (1971).

18. V. A. Komissaruk, Tr. Voenno-Vozdushn. Inzh. Akad. im. Zhukovskogo, No. 1301, 124 (1971).

19. L. A. Vasil'ev and I. V. Ershov, *A Grating Interferometer* (Mashinostroenie, Moscow, 1976) [in Russian].

Translated by A. Chikishev

OPTICS,
QUANTUM ELECTRONICS

Photoemission Pulsed Source of Wide-Band Directional Electromagnetic Radiation

Yu. N. Lazarev, P. V. Petrov, and Yu. G. Syrtsova

Federal State Organization Russian Research Nuclear Center,
All-Russia Institute of Technical Physics, Snezhinsk, Chelyabinsk Oblast, 456770 Russia

e-mail: pvpetrov@snezhinsk.ru

Abstract—Generation of wide-band directional electromagnetic radiation arising when the pulsed X radiation front strikes the photocathode of a planar diode at an angle is analyzed. The results of numerical simulation are compared with the experimental data obtained with the Iskra-5 setup, which is used for generation of a laser plasma as an X-ray source. © *2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The existence of faster-than-light velocities and faster-than-light sources (FTLSs) of electromagnetic radiation has been known for a long time. Generally speaking, these sources move with a velocity exceeding the phase velocity of light: $v_{ph} > c//n$, where n is the refractive index of the medium. It is a matter of common knowledge that such sources exist in media with $n > 1$. However, their existence in a vacuum is much less known [1]. Of course, the case in point is the motion not of a point particle but of some effective charge distribution (macroscopic charge) with a velocity higher than the velocity of light in a vacuum. Such a distribution may be achieved using slower-than-light motions of real charges [1, 2]. Theoretical analysis of the radiation from FTLSs may be performed using the apparatus of the conventional field theory [3].

The object of our study is FTLSs that arise due to electrons emitted from the vacuum–medium interface where the emission front propagates with a velocity exceeding velocity of light c. A simple source of this kind is formed when a flat metal surface is irradiated by a ribbon beam of ionizing radiation (Fig. 1a) [4–6]. In this case, the phase velocity of the emission front is

$$v_{ph} = c/\sin(\theta), \tag{1}$$

where θ is the angle of incidence of the radiation.

The radiation due to the emission current propagating with constant velocity $v_{ph} > c$ along the screen by no means boils down to the Vavilov–Cherenkov effect [1], although the resulting electromagnetic wave leaves the interface at an angle $\theta_w = \arcsin(c/v_{ph})$, which is typical of Vavilov–Cherenkov radiation and equals the angle of incidence. At distance R far away from the source, the amplitude of the resulting electromagnetic

wave is proportional to the second-order time derivative of the dipole moment surface density, which varies as the kinetic energy ε of the electrons knocked out and surface area S of the radiator [4, 5]:

$$E^w \sim H^w \sim \frac{\ddot{P}S}{c^2 R}, \quad \ddot{P} = \left|\frac{d^2\mathbf{P}}{dt^2}\right| \sim \varepsilon,$$
$$\mathbf{P}(\mathbf{r}, t) = \int dV'(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r}', t). \tag{2}$$

Formula (2) implies that the intensity and total energy of the electromagnetic radiation grow with the energy of the electrons emitted. Estimates [5, 6] show that the energy and intensity of the radiation are of practical interest if the electrons emitted have an energy above 10 keV. In the laboratory conditions, a faster-than-light current pulse with a high electron energy can be produced if the process of electron generation is separated from the formation of the faster-than-light pulse. In this case, optical or ionizing radiation (IR) that may cause the electron emission is used only for generating electrons with an energy as low as possible, while the final energy on the order of 100 keV is gained when the electrons are accelerated in an external electric field. A simple design of an FTLS includes a planar accelerating diode with a grid anode that is exposed to an inclined beam of ionizing radiation [5–7]. The IR front incident on the photocathode generates a faster-than-light current pulse of the electrons emitted. Then, the electrons are accelerated in the electrode gap, pass through the grid anode, and generate another faster-than-light current pulse above the anode (but this pulse comprises fast electrons!). It should be noted, however, that the generation and acceleration processes alter the space–time electron distribution, so that the time dependence of the current at the anode and cathode will
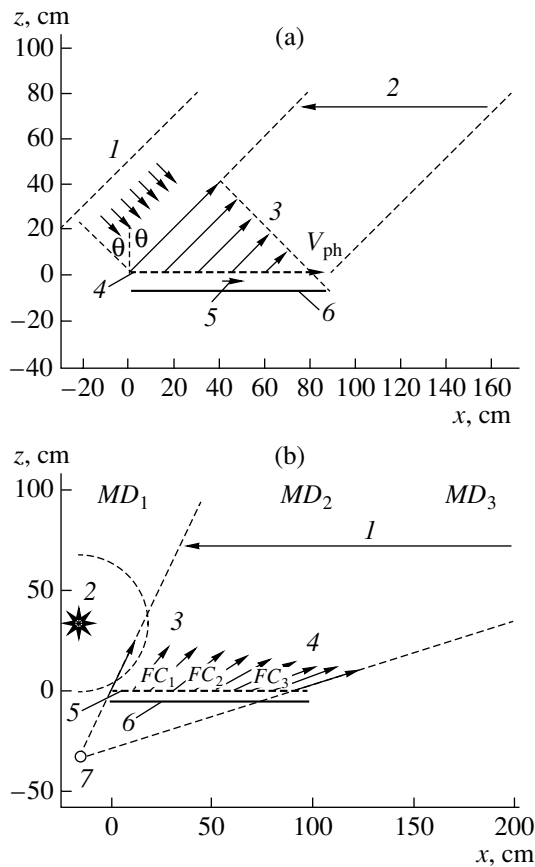
Fig. 1. Basic scheme of EMR generation by an FTLS. (a) *1*, ribbon X-ray beam; *2*, region of faster-than-light radiation; *3*, directional electromagnetic wave front; *4*, anode; *5*, waveguide electromagnetic wave; and *6*, metallic cathode. (b) *1*, region of faster-than-light radiation; *2*, point source of X rays; *3*, spherical front of X radiation; *4*, direction of EMR propagation; *5*, anode; *6*, metallic cathode; and *7*, X-ray source image.

be different. Thus, a dipole layer radiating in phase is formed above the grid anode.

The design considered offers a number of attractive properties that are a direct consequence of the fact that phasing of electrons is accomplished by the faster-than-light pump pulse moving over the surface [5–7]. These properties are as follows: (i) the radiation is coherent and directional; that is, radiated energy Q is proportional to emitting surface area S; (ii) the videopulse radiated is short: its duration is roughly expressed as

$$T_0 \approx \frac{L}{c}\sqrt{(\gamma + 1)/(\gamma - 1)},$$

where

$$\gamma = 1 + \frac{\varepsilon}{m_e c^2},$$

L is the electrode-to-electrode spacing, and m_e is the electron mass; (iii) the process is highly efficient: theo-

retically, the radiated energy constitutes a fraction $\sqrt{(\gamma - 1)/(\gamma + 1)}$ of the total electrostatic energy stored in the capacitor; and (iv) the efficiency of the emission increases in inverse proportion to the wavelength λ of the resulting radiation. The last-named property is unique: most microwave generators exhibit the opposite tendency.

Varying the parameters of the radiator (diode), as well as using carious sources of optical radiation and various power supplies for the diode, one can design a variety of devices generating a microwave pulse.

The feature of such a diode-type FTLS is the generation of two electromagnetic waves (hereafter referred to as directional and waveguide), which interact with the electrons emitted, thereby defining the anode current and parameters of the electron dipole moment above the anode. The directional wave propagates in the free half-space over the anode in the direction that is "specular" relative to the direction of the incident IR ($z > 0$, Fig. 1a). The parameters of this wave depend on the rate of change of the parameters of the dipole layer made up by the electrons accelerated in the diode (i.e., on $\dot{P}$ and $\ddot{P}$; see (2)). It may be assumed that this wave insignificantly affects the current near the cathode because of the screening effect of the metallic grid anode. The faster-than-light current pulse due to the electron motion in the electrode gap also excites an electromagnetic wave that propagates in the planar waveguide formed by the cathode and anode. The field of this waveguide electromagnetic wave adds up with the field of the space charge of the electrons emitted, effectively reducing the accelerating field in the diode and, thus, decreasing the ultimate current density at the anode below the limiting stationary current density. It was shown analytically [7] that the parameters of the elementary FTLS depend on the rate of rise of the electron emission current from the cathode and on the parameters of the diode (the voltage across and the width of the accelerating gap).

In the experiments on EMR generation with an FTLS that were carried in the All-Russia Institute of Technical Physics (ARITP) [8], the emission of electrons was initiated by a short X-ray pulse from a point laser plasma source. The pulse fell at an angle to the cathode of a planar diode. The experiments were aimed at (i) designing FTLSs, (ii) studying their characteristics, and (iii) comparing the experimental data with the theory.

In this work, we report the results of mathematical simulation of EMR generation that was made in a statement closely approximating the experimental conditions used in the ARITP and compare the analytical and experimental data.

EXPERIMENTAL

The experimental scheme was selected based on the data obtained in [5, 6], where faster-than-light current pulses and directional EMR were generated using an accelerating diode with a transparent anode. The electron emission was initiated by an obliquely incident ribbon beam of IR.

Unfortunately, the experimental equipment currently available is incapable of forming a ribbon beam of duration and intensity sufficient for generating an emission current pulse propagating with a constant faster-than-light velocity (see (1)) along the photocathode. A spherical front of IR from a point source is much easier to obtain, and this shape was used in the experiments [8]. In those experiments, the electron emission was initiated by soft X rays leaving the plasma produced by focusing the subnanosecond radiation (wavelength $\lambda = 1.315$ μm) from the Iskra-5 setup [9] onto a gold target. The focal spot was ≈ 1 mm in diameter. This suggests that the IR source is point and the front of the IR is spherical.

Use of a point source to irradiate a conducting plate is an essential point that differentiates the experiments described from those in [5–7]. Unlike a flat-front FTLS, which generates a current pulse with a velocity that is constant throughout the emitting surface, the use of the spherical front causes the angle of incidence of the radiation to vary (increase) and the faster-than-light velocity of the emission current along the photocathode to decrease. Since the direction of emission of the electromagnetic wave depends on the faster-than-light velocity value (and this velocity, in turn, is related to the angle of incidence of the X rays), we obtain a divergent EMR beam (instead of a ribbon beam for the case of the flat X-ray front, Fig. 1a). The parameters of this divergent beam depend on the mutual arrangement of the point X-ray source and plate irradiated, as is schematically (without regard to diffraction divergence) shown in Fig. 1b.

The essence of our experiments is shown in Fig. 1b. We used a 3-m-long echo-free vacuum chamber of diameter 2 m with the microwave-absorbing inner surface. The chamber enclosed an electromagnetic radiator to be tested (a plane capacitor with a transparent plate (anode)), an X-ray source, X-ray detectors, accelerated-electron current detectors, and EMR detectors.

The diode represented a plane capacitor with a polished aluminum cathode measuring 60×850 mm and an anode (nickel grid with a mesh size of 2 mm). The anode transmitted about 80% of incident light. The source was mounted at a height of 34 cm above the photocathode and was 40 cm distant from its nearest end. The gap between the electrodes was within 20 mm. The grid was under an accelerating potential $\varphi_0 \approx 80$ kV.

To protect the inner surface of the chamber and the bodies of the electromagnetic field detectors against the direct action of X rays, the source was sheathed in a metallic screen with holes through which the incident

radiation was directed toward the cathode and X-ray detectors.

The local values of accelerated-electron current density at the outer surface of the anode were measured with three Faraday cups FC_1, FC_2, and FC_3. By means of inductive magnetic-field detectors, we measured the amplitude–time parameters and took the directivity diagram of the EMR. The inductive detector was a frame antenna in the form of a wire turn 10 mm in diameter placed on a metallic plate. The magnetic detectors (MD_1–MD_3) were aligned with a line parallel to the cylinder generatrix near the microwave-absorbing coating. MD_3 was placed in the area where the characteristic radiation of the FTLS was expected to be the highest intensity, MD_1 was placed outside the area of radiation, and MD_2 was in between (Fig. 1b). The distance from MD_3, MD_2, and MD_1 to the center of the capacitor was 200, 120, and 120 cm, respectively.

The output signals from the EMR detectors and the electron currents were measured using high-speed oscillographic recorders with bandwidths of 5 and 7.5 GHz. The transient time in the recording channels was no more than 100 ps. The accuracy of finding the time parameters was 50 ps or higher. The relative error of measuring the current amplitudes was less than 20%. The field amplitudes were measured accurate to 25%. The error of the relative measurements did not exceed 10%.

EXPERIMENTAL RESULTS

In the experiments [8], laser energy E delivered to the target was varied between 0.3 and 0.8 kJ; pulse duration $\tau_{0.5}$, between 0.3 and 0.5 ns; and irradiation fluence Q, between 10^{14} and 10^{15} W/cm^2. The duration of X-ray pulses for photons of energy $\varepsilon \sim 0.45$ keV was about 0.7 ns at a pulse rise time of ~0.3 ns. For the same rise time, the duration of a pulse of "complete" plasma radiation was between 2 and 3 ns. The effective temperature of the X-ray source spectrum was about 50 eV.

As follows from the readings of the X-ray detectors, the X-ray intensity at the near and far (relative to the source) ends of the cathode differed by a factor of five to six.

The high fluence of X-rays makes it possible to generate an emission pulse of current density J_{ca} varying from several tens to several hundreds of A/cm^2 at the cathode end nearest to the X-ray source with pulse rise time T_{rc} of about 0.3 ns. With such a high rate of electron emission, the time it takes for the capacitor to be discharged completely at an initial voltage of 80 kV and a electron current density of 100 A/cm^2 is

$$T_m \approx \sqrt{\varphi_0 T_{rc}/2\pi L J_{ca}} \approx 0.2 \text{ ns.} \qquad (3)$$

This indicates the formation of an FTLS at the front of the pump pulse. Under these conditions, the parameter responsible for its properties is not the amplitude
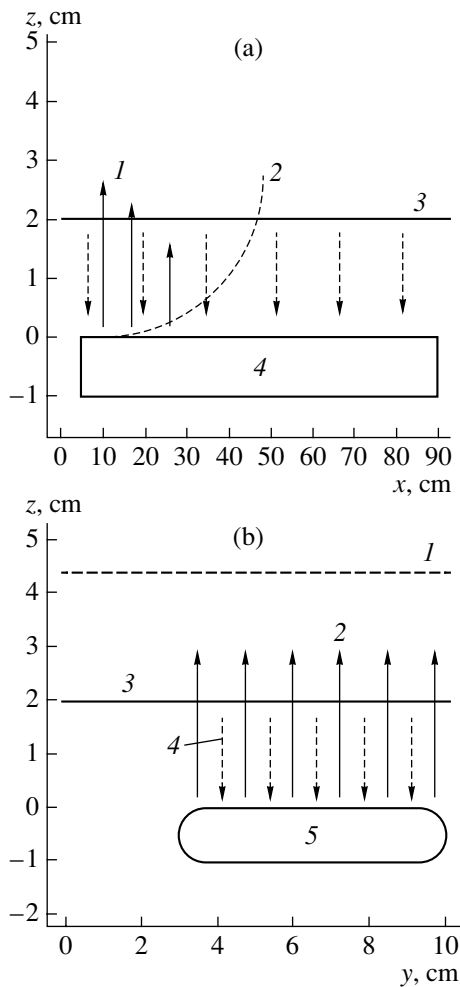
Fig. 2. Schemes of capacitor irradiation in the FTLS simulation (2D geometry). (a) *1*, electrons; *2*, cylindrical X-ray beam; *3*, anode; and *4*, cathode. (b) *1*, ribbon X-ray beam; *2*, emitted electrons; *3*, anode; *4*, accelerating electric field; and *5*, cathode.

value of the emission current at the cathode (or the maximal fluence of X-rays), which was measured fairly accurately in the experiments, but the rise time of the electron emission current (or the rise time of the X-ray pulse), which is much more difficult to measure.

From the measurements of the amplitude and rise time of the anode current, as well as of the electromagnetic field amplitudes, we can draw the following conclusions.

(1) The faster-than-light current pulse and EMR are generated under the saturation conditions: the parameters of the current pulse at the anode and the parameters of the electromagnetic wave emitted (amplitude and rise time) are virtually independent of the amplitude of the emission current at the cathode (or of the X-ray fluence).

(2) The measured values of the anode current are lower than the ultimate value of the stationary current:

$J_0 = \varphi_0^{3/2}/9\pi L^2 \sqrt{2e/m} \simeq 13.2 \text{ A/cm}^2$. This is consistent with the idea that the electron current in FTLSs is dynamically limited by the field of the EMR generated [7].

(3) As follows from the readings of the magnetic detectors placed in different areas of the chamber, the characteristic radiation of the FTLS is directional and basically corresponds to the scheme shown in Fig. 1b.

SIMULATION OF FASTER-THAN-LIGHT SOURCE PARAMETERS

Since the EMR generation, the formation of the dipole layer, and the motion of electrons in the electrode gap are intimately related processes, they should be considered in combination. The basic means for theoretical investigation of such systems is numerical simulation based on the self-consistent solution of the Maxwell and Vlasov equations

$$\frac{\partial f_e}{\partial t} + \mathbf{v}\nabla f_e - e\left\{\mathbf{E}^w + \mathbf{E}_0 + \left[\frac{\mathbf{v}}{c}, \mathbf{H}^w\right]\right\}\frac{\partial f_e}{\partial \mathbf{p}} = 0; \quad (4)$$

$$\mathbf{j}(\mathbf{r}, t) = -e\int d^3p \cdot \mathbf{v}f_e(\mathbf{r}, \mathbf{p}, t);$$

$$\text{curl}(\mathbf{E}^w) = -\frac{1}{c}\frac{\partial \mathbf{H}^w}{\partial t}; \quad \text{curl}(\mathbf{H}^w) = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{E}^w}{\partial t}. \quad (5)$$

Here, f_e is the electron distribution function; e is the electron charge; $\mathbf{p}$ and $\mathbf{v}$ are the momentum and velocity of an electron, respectively; $\mathbf{E}^w$ and $\mathbf{H}^w$ are the strengths of the electric and magnetic fields, respectively, that are generated by the electrons emitted; and $\mathbf{E}_0$ is the electric field strength in the accelerating diode.

The simulation was carried out in two stages. At the early stage, the EMC2D two-dimensional program [10] was used. In terms of this program, the Vlasov equation is solved by the particle-in-mesh method [11]; the Maxwell equations, by the finite-difference method based on the cross scheme [12]. With this scheme, the formation of a faster-than-light current pulse, $\{j_x^{2D}, j_z^{2D}\}$, and the generation of the EMR, $\{E_x, E_z, H_y\}$, were self-consistently simulated in the Cartesian system. The system geometry used in the simulation is shown in Fig. 2. Two irradiation schemes were employed. In the first case, the capacitor, which was assumed to be infinitely long in the $0y$ direction, was irradiated by a cylindrical X-ray beam (Fig. 2a). Here, the space–time distribution of the electric current, $\{j_y^{2D}, j_z^{2D}\}$, and the derivative $\partial P_z^{2D}(t, x)/\partial t$ of the dipole moment along the $0x$ axis (along the larger side of the anode) were determined and the electromagnetic fields at the detectors were calculated. In the second scheme, the capacitor, which was assumed to be infinitely long in the $0x$ direction, was irradiated by a ribbon X-ray beam (Fig. 2b).

Here, the space–time distribution of the electric current, $\{j_y^{2D}, j_z^{2D}\}$, and the derivative $\partial P_z^{2D}(t, y)/\partial t$ of the dipole moment along the $0y$ axis (along the smaller side of the anode) were determined.

At the second stage (3D geometry of simulation), the electromagnetic fields at the detectors were simulated for the radiator with the actual position and dimensions of the diode. The simulation was based on the GEMC 3D program, which makes it possible to solve Maxwell equations (5) by the finite-difference method [12] in the Cartesian system for the six electromagnetic field components E_x, E_y, E_z, H_x, H_y, and H_z given the current found at the first stage. Since the complete space–time distribution of the electric current found with the EMC2D program is difficult to specify, we used the dipole approximation

$$j_{z,S}^{3D}(t, x, y) = \dot{P}_z^{2D}(t, x)\frac{\dot{P}_z^{2D}(t, y)}{\dot{P}_z^{2D}(t, y = y_0)}K_{2D},$$

$$\dot{P}_z^{2D}(t, \alpha) = \frac{\partial P_z^{2D}(t, \alpha)}{\partial t} = \int_0^\infty dz' j_z^{2D}(t, \alpha, z'), \quad (6)$$

$$\alpha = x, y; \quad K_{2D} = \frac{\dot{P}_z^{2D}(t, y = y_0)}{\dot{P}_z^{1D}(t)}.$$

Here, $j_{z,S}^{3D}$ is the electron current surface density obtained by the 3D simulation; $j_z^{2D}(t, \alpha, z)$ and $\dot{P}_z^{2D}(t, \alpha, z)$ are the electron current density and the derivative of the dipole moment density, respectively, that were obtained by the 2D simulation; and K_{2D} is the coefficient taking into account the fact that the derivative $\dot{P}_z^{2D}(t, x)$ of the dipole moment density changes in going from the infinite (along the $0y$ direction) to finite plate.

In our case, dipole approximation (6) is valid, since we are interested in the early stage of generation, when the size $\Delta z \sim L$ of the radiating layer is smaller than, or comparable to, the wavelength $\lambda \sim 2\pi c/v_0 L$ ($v_0 \lesssim 0.5c$) of the resulting radiation.

In the calculations, the initial and boundary conditions were set as follows. (1) For a given geometry and voltage of the accelerating diode, the initial values of the electric field were calculated by the finite-element method with the MATHLAB package [13] and then converted to those obtained with the EMC2D finite-difference grid; (2) the initial values of the magnetic field were taken to be equal to zero; (3) escape conditions for the particles and electromagnetic waves at the boundary of the domain of simulation were set; and (4) an electron emission source was set on the cathode of the

accelerating diode with allowance for X radiation attenuation and a time delay associated with a finite propagation velocity of the IR spherical front along the cathode:

$$f(t, z = 0, t, \mathbf{p}) = \frac{J_c(t - x/v_{ph}(x))}{4\pi((x - x_s)^2 + z_s^2)}$$

$$\times \eta(t - x/v_{ph}(x))\frac{2}{\sqrt{\pi}p_0}\exp(-p_z^2/p_0^2), \quad (7)$$

where p_0 is the mean momentum of the electrons emitted, $J_c(t)$ is the time-varying emission current, and $\{x_s, z_s\}$ are the coordinates of the X-ray source.

A feature of FTLS simulation is that not only the time dependence of the emission current but also its derivative $dJ_c(t)/dt$ (the rise time of the current at the cathode) should be specified with a reasonable accuracy. It is this derivative that defines the parameters of the resulting radiation under the saturation conditions [7]. In the experiments [8], the emission current was so high that direct measurement of the time parameters of the current pulses was impossible in the linear operating mode of the detectors. Specifically, this is also true for the rise time of the signal, $J_c'(t)$, for $0 < t < 2$ ns, which makes experimental data interpretation uncertain.

In view of the aforesaid, it becomes clear that, for function $J_c(t)$ to be adequately set using experimental data, it must be more accurately defined for short time intervals. This can be done in a number of ways, for example, by averaging all available experimental dependences. Alternatively, of these dependences, one can choose a function with a certain value of $J_c'(t \longrightarrow 0)$ based on the calculated data for the X radiation at the exit from a laser plasma source. Irrespective of the approach to selecting $J_c(t)$, this function must be in fairly good agreement with the available experimental data for the parameters of the anode current. The agreement may be checked if it is taken into account that the rise time of the electron emission current density is in one-to-one correspondence with the amplitude and rise time of the anode current density. When checking, we used anode current amplitudes measured by a Faraday cup FC_1.

Figures 3 and 4 show the density and rise time (at a level of 0.1–0.9 of the peak value) of the anode current that were calculated at FC_1 (roughly 20.5 mm away from the near edge of the anode) versus the emission current amplitude at the anode for various time dependences of the emission current (these dependences were recorded in a number of experiments). As the emission current increases, the anode current does not decrease and the rise time shortens.
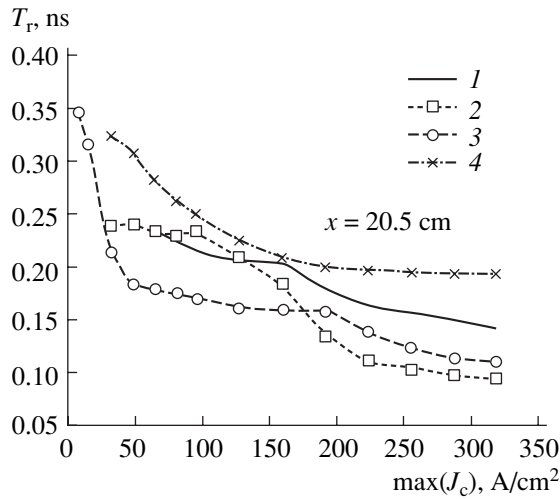
Fig. 3. Anode current rise time vs. emission current amplitude at the cathode for various emission current pulse shapes: *1*, averaged shape of $J_c(t)$; *2*, first shape of $J_c(t)$; *3*, second shape of $J_c(t)$; and *4*, third shape of $J_c(t)$.
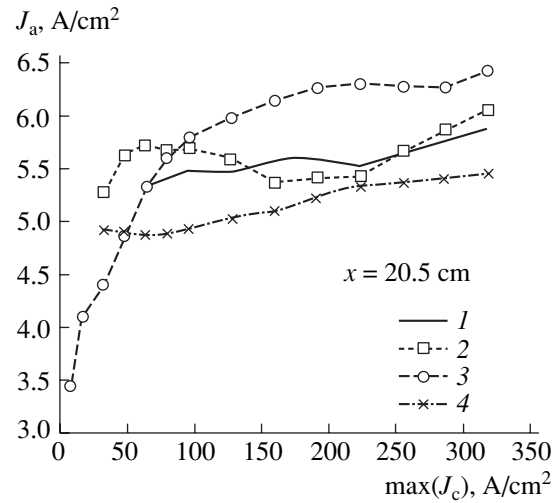

Fig. 4. Anode current amplitude vs. emission current amplitude at the cathode for various emission current pulse shapes. *1–4* are the same as in Fig. 3.
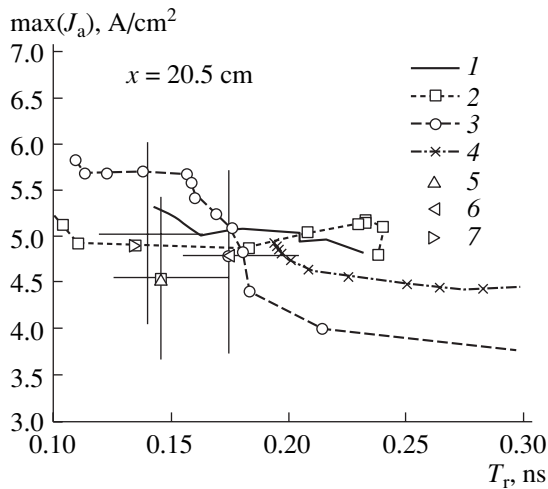

Fig. 5. Anode current amplitude vs. anode current rise time dependences obtained experimentally and calculated for various emission current pulse shapes. (*1–4*) the same as in Fig. 3; (*5–7*) experiments 2409, 1804, and 1304, respectively.
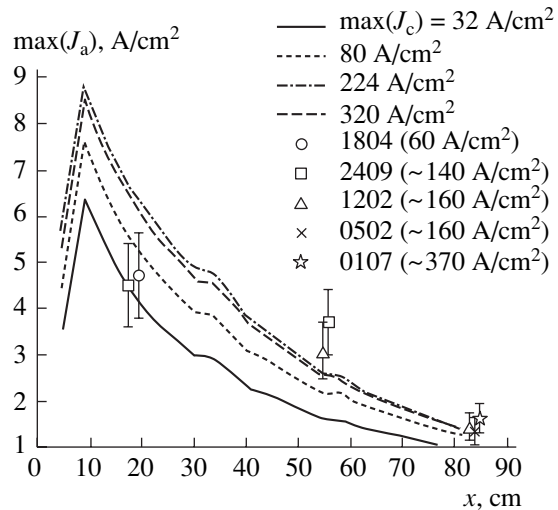
Fig. 6. Spatial distribution of the anode current amplitude along the cathode for various emission current amplitudes at the cathode.

Let us use the data on the anode current amplitude and rise time for selecting $J_c(t)$. Figure 5 demonstrates the anode current density versus rise time dependences obtained by the simulation and in the experiments. At first glance, a slight uncertainty in the emission current pulse shape causes noticeable uncertainties in the anode current amplitude (up to 1 A/cm²) and rise time (to 0.1 ns). However, it may be said that the first, second, and averaged shapes of the emission current pulse agree with the anode current measurements within the accuracy of anode current measurement. Subsequently, when simulating the emission current at the cathode, we used the first shape of the emission current pulse.

EXPERIMENTAL DATA VERSUS RESULTS OF SIMULATION

The parameters of the faster-than-light current pulse (waveform and amplitude variation along the cathode), which propagates over the anode, to the greatest extent define the space–time distribution of the resulting EMR. Figures 6 and 7 show the experimental and analytical distributions of the anode current density and rise time along the radiator. The experimental data and the results of calculation are seen to be in good agreement. It should also be noted that the emission current amplitude at the cathode varies in the interval 30–
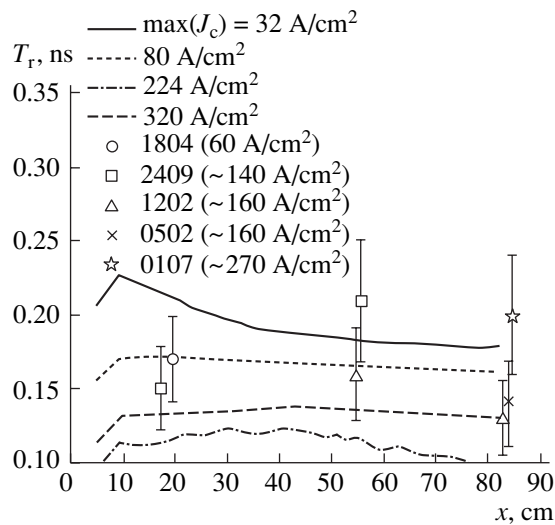
Fig. 7. Spatial distribution of the anode current rise time along the cathode for various emission current amplitudes at the cathode.
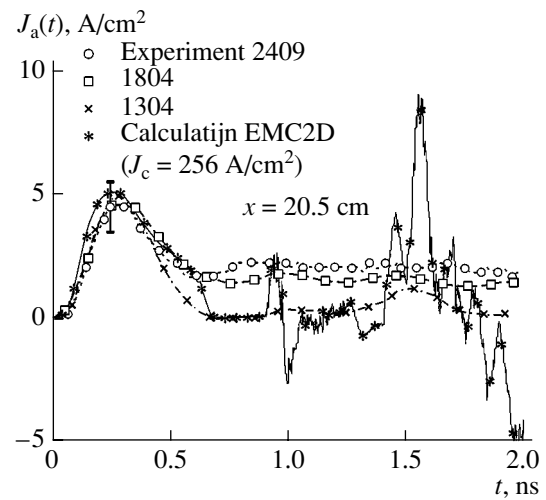

Fig. 8. Time dependences of the accelerated electron current over the anode at the point of FC_1 location (the edge of the capacitor that is nearest to the X radiation source).
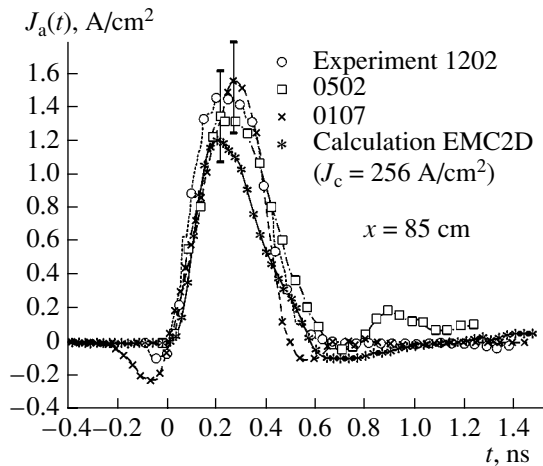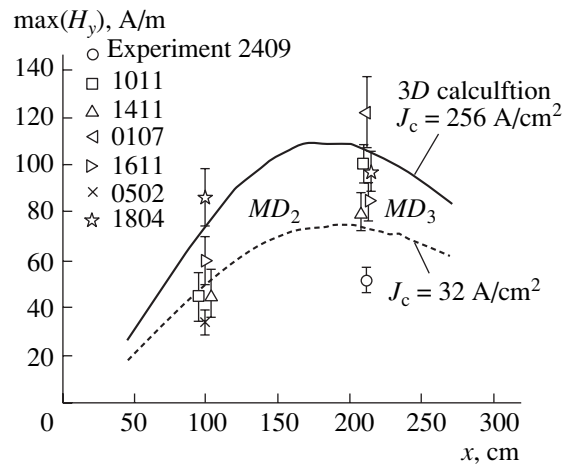

Fig. 9. Time dependences of the accelerated electron current over the anode at the point of FC_1 location (the far end of the capacitor).

Fig. 10. Spatial distribution of the maximal magnetic field along the $0x$ axis, which is parallel to the larger side of the capacitor and passes through the points of magnetic detector location.

300 A/cm², which also agrees with the experiment. The anode current rapidly drops with distance to the source, with the rise time remaining almost unchanged (Fig. 7) but the pulse shape varying noticeably (Figs. 8, 9). It was shown [7] that this effect is related to the dynamic limitation of the anode current by the field of the waveguide mode, which is generated by a faster-than-light current pulse due to the electrons accelerated in the electrode gap. The net field from the waveguide electromagnetic wave and space charge builds up much faster and is of much greater importance than the space charge field alone. The net field cuts time T_m, in which electrons are injected from the cathode and then reach the cathode, and also dynamically decreases the accelerating field in the electrode gap.

As a result, the anode current and amplitudes of the derivatives of the dipole moments also decrease. Since the FTLS at the X radiation front forms under the saturation conditions, the irradiation nonuniformity due to a decrease in the X ray fluence has no time to show up against the background of the effect of current limitation by the waveguide mode and, hence, insignificantly influences the generation of the faster-than-light current pulse and EMR.

These effects are distinctly seen in Figs. 8 and 9, which plot the time dependences of the anode current that are measured at different points along the $0x$ axis and obtained analytically (by FTLS simulation). It should be noted that, since the time resolution of the measuring channel is comparable to the microwave
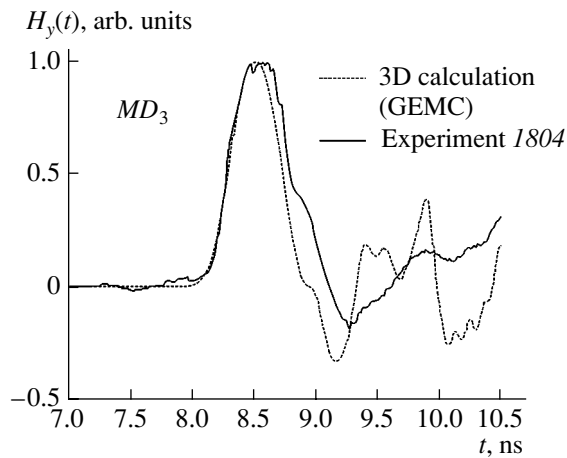
$H_y(t)$, arb. units

Fig. 11. Experimental and calculated time dependences of the magnetic field at the point of MD_3 location.

component of the current pulse, the experimental signals may be smoother than those obtained by the calculation.

The simulation shows that, for the given intensities of the X radiation from the laser plasma source, the anode current rapidly tends to an asymptotic value (in terms of the emission current at the cathode) and its peak value at the point of FC_1 location is bound to be no more than 6.5 A/cm^2. This statement was confirmed experimentally, since an anode current density higher than 5 A/cm^2 was observed in none of the experiments.

One of our main goals was to record the EMR from the FTLS and confirm that it is directional and short. The magnetic fields were recorded in the far-field zone of the radiator, where the field amplitudes depend primarily on the second-order derivative of the dipole moment. This derivative is proportional to the electron energy and initial voltage across the capacitor: $\ddot{P} \sim \varepsilon \sim \varphi_0$. Since there is an uncertainty both in the pulse shape and in the emission current amplitude at the cathode (because of the experimental error involved in the experimental values of the anode current and its rise time; see Figs. 3, 4), one can indicate only the interval into which the values measured may fall when simulating the magnetic field at the points where detectors MD_2 and MD_3 are located. Figure 10 shows the distribution of the maximal magnetic field along the $0x$ axis that was simulated for those coordinates x and y corresponding to the positions of the magnetic detectors. The results of simulation are in good agreement with the measured data both in magnitude and in pulse shape (Fig. 11). In the wave zone of the radiator, the resulting EMP represents a videopulse of duration less than 2 ns (this duration depends on the characteristic discharge time of the capacitor). With the rise times coincident, the analytical and calculated pulses somewhat differ in FWHM (by 0.1–0.2 ns). This discrepancy is most probably related to the fact that, in the 3D simulation of the

electromagnetic fields, the electron currents were specified by the time functions obtained by the self-consistent simulation in the 2D geometry, which ignores the finite length of the capacitor in the $0y$ direction. The calculation of the current distribution in the direction transverse to the direction of the faster-than-light current (Fig. 2b) indicates that not only does the anode current amplitude rise by 20–30%, but also the pulse becomes slightly (by 10–15%) longer.

CONCLUSIONS

We performed mathematical simulation of the EMR generation by a faster-than-light emission current source, which was implemented with the Iskra-5 setup in the ARITP [8]. The simulation shows that the characteristic EMR is directional and lasts for a short time. Also, the resulting EMR induces the dynamic limitation of the anode current density.

In the calculation, we determined the anode current values that are ultimate in such a system. The space–time distribution of the emitting dipole layer was obtained. It is shown that the anode current amplitude and the dipole moment density rapidly decay along the capacitor and tend to their asymptotic values as the length of the emitting area increases. This is associated with the current limitation effect and also with the influence of the resulting EMR on the electron motion in the accelerating gap.

Generally, we may ascertain that the physicomathematical model of an FTLS gives a good fit to the experimental space–time distribution of the anode current and electromagnetic fields. Hence, the concepts this model relies upon [1, 4–7] are valid.

REFERENCES

1. V. L. Ginzburg, *Theoretical Physics and Astrophysics* (Pergamon, Oxford, 1979; Nauka, Moscow, 1987).

2. B. M. Bolotovskiĭ and V. L. Ginzburg, Usp. Fiz. Nauk **106**, 577 (1972) [Sov. Phys. Usp. **15**, 184 (1972)].

3. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 2: *The Classical Theory of Fields* (Nauka, Moscow, 1988; Pergamon, Oxford, 1975).

4. N. J. Carron and C. L. Longmire, IEEE Trans. Nucl. Sci. **23**, 1897 (1976).

5. Yu. N. Lazarev and P. V. Petrov, Zh. Éksp. Teor. Fiz. **115**, 1689 (1999) [JETP **88**, 926 (1999)].

6. Yu. N. Lazarev and P. V. Petrov, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 625 (1994) [JETP Lett. **60**, 634 (1994)].

7. Yu. N. Lazarev, P. V. Petrov, and Yu. G. Syrtsova, Fiz. Plazmy **29**, 491 (2003) [Plasma Phys. Rep. **29**, 491 (2003)].

8. A. V. Bessarab, A. V. Kunin, S. P. Martynenko, *et al.*, Tr. RFYaTs–VNIIÉF, No. 1, 518 (2001).

9. V. I. Annenkov, V. A. Bagretsov, V. G. Bezuglov, *et al.*, Kvantovaya Élektron. (Moscow) **18**, 536 (1991).

10. E. V. Diyankova and P. V. Petrov, Preprint No. 99, VNI-ITF (Russia Landau Institute for Theoretical Physics, Russian Academy of Sciences, 1996).

11. R. Hockney and J. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1984; Mir, Moscow, 1987).

12. J. W. Eastwood, Comput. Phys. Commun. **64**, 252 (1991).

13. *Partial Differential Equation Toolbox User's Guide* (Mathworks, 1997).

Translated by V. Isaakyan

OPTICS,
QUANTUM ELECTRONICS

Nonlinear Mid-IR Radiation in Two-Frequency Semiconductor Lasers with a Corrugated Waveguide

V. Ya. Aleshkin*, A. A. Afonenko**, and A. A. Dubinov*

* Institute of Physics of Microstructures, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia
e-mail: sanya@ipm.sci_nnov.ru
** Belarussian State University, Minsk, 220050 Belarus
Received February 6, 2004

Abstract—The nonlinear generation of a difference frequency mode in an injection quantum-well semiconductor laser is considered. A laser based on the InGaAs/GaAs/InGaP heterostructure is proposed, which generates two modes in the 1-μm range and the difference mode in a corrugated waveguide in the range from 10 to 20 μm. It is shown that the power of the difference mode produced by a laser with a 100-μm-wide waveguide in the mid-IR range at room temperature can be as high as a few microwatts if the power of the short-wave modes is 10 W. © *2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Although small-size terahertz and multiterahertz sources of radiation are presently in great demand in various applications, only few types of such sources have been developed. Semiconductor lasers of traditional design are only capable of lasing in the near-infrared (IR) and visible ranges, because nonradiative Auger recombination plays a significant role in narrow-gap semiconductors. Presently, the most significant advances are observed in the development of quantum cascade lasers [1, 2]. However, the extremely complex band structure of cascade lasers and stringent requirements imposed on their parameters hinder their wide use. Far-IR semiconductor lasers based on *p*-type germanium [3, 4] operate only at cryogenic temperatures. An alternative approach to the development of mid- and far-IR lasers operating at room temperature is to use nonlinear effects. Previously, to produce a difference mode in a laser that outputs two short-wave modes, it was proposed to use electron nonlinearity in a quantum well containing three levels [5] or nonlinear properties of semiconductor materials in the active region [6, 7].

The main difficulty in achieving efficient nonlinear generation consists in the necessity to meet the phase-matching condition, because, due to the normal dispersion of the refractive index, the phase velocity of the wave of nonlinear polarization, as a rule, proves to be lower than the phase velocity of the difference mode. It was shown in [6] that the phase-matching condition could be met when the fundamental short-wave mode at a frequency ω_1 and a side mode at a frequency $\omega_2 > \omega_1$ are used. In this case, high-frequency modes at about 1-μm with a power of 10 W produce a 100-μW difference mode at a wavelength of about 10 μm and an absorption coefficient of about 10 cm^{-1}.

A disadvantage of the design proposed in [6] is a small overlap coefficient for the wave of nonlinear polarization at the difference frequency and the waveguide mode excited. This coefficient is small for two reasons. First, the nonlinear polarization at the difference frequency is proportional to a product of almost orthogonal modes, which changes the polarization sign in the direction perpendicular to the structure layers on the scale of the width of the wave-guiding layer for high-frequency modes. Second, the scale of the spatial variation of the difference mode in this direction is mush larger than the scale of variation of the polarization wave. In addition, the parameters of the dielectric waveguide proposed in [6] are very sensitive to the geometry of the structure; hence, even small errors in the layer thicknesses may decrease the laser power by several orders of magnitude, which makes this design very difficult to implement.

An alternative approach to providing phase matching by using a plasma waveguide to slow the difference harmonic was proposed in [7]. However, as calculation showed, this structure operates satisfactorily only for the difference mode is in the far-IR range (where the semiconductor permittivity exhibits anomalous dispersion) and is hardly applicable to the mid-IR range.

In this study, we propose a new technique to ensure the phase matching, based on using modes of a corrugated waveguide, which has long been used to slow electromagnetic waves in microwave electronics [8].

The advantages of this technique are as follows. First, it offers a comparatively easy way to control the phase velocity at the difference frequency by changing the corrugation period and depth. The second advantage of the proposed technique is that it uses two high-frequency fundamental modes at frequencies ω_1 and ω_2 to excite the low-frequency difference mode; therefore,

the overlap coefficient for the difference frequency is not small (the high-frequency modes are not orthogonal). Our calculations showed that, when the phase-matching condition is met and the powers of the high-frequency modes are 10 W at about 1 μm, the power of the difference mode may be about several microwatts in the wavelength range from 10 to 20 μm for a laser with a 100-μm-wide waveguide.

Fig. 1. Structure proposed to generate the difference frequency (layer numbers correspond to those used in table).

CALCULATION OF THE POWER OF THE DIFFERENCE MODE IN THE CORRUGATED WAVEGUIDE

Consider the structure in Fig. 1. Here, h is the corrugation depth, D is the corrugation period, and L is the length of the region occupied by the semiconductor and surrounded by metal from its three sides. If the semiconductor laser structure is grown on the (001) plane and the high-frequency modes are TE-polarized and propagate in the [110] direction, the nonlinear polarization in GaAs is perpendicular to the layer plane and excites the TM mode at the difference frequency [6].

We will calculate the characteristics of the corrugated waveguide at the difference frequency by the mode-matching technique [8], which divides the corrugated waveguide into two regions: the space of resonators ($z > 0$) and the interaction space ($z < 0$). The H_y magnetic field component of the TM wave can be approximated in the interaction space by a superposition of spatial harmonics $H_{ym}(z)$; in the resonators, by a superposition of partial fields $H_{yq}(z)$, which are standing waves in the propagation direction of the wave amplified, i.e., they do not contribute to the energy transfer:

$$H_y(x, z) = \sum_{m = -\infty}^{+\infty} H_{ym}(z) \exp(ik_m - i\omega t), \quad z < 0, \quad (1)$$

$$H_y(x, z) = \sum_{q = 0}^{\infty} H_{yq}(z) f_q(x) \exp(-i\omega t), \quad z > 0, \quad (2)$$

where $k_m = k_x + 2\pi m/D$; $-\infty < m < \infty$; and $f_q(x)$ are the resonator's eigenfunctions ($0 \leq q < \infty$), whose phase advance per corrugation period D equals the phase advance of the wave outside the resonators.

We used the single-wave approximation, which takes into account only one spatial harmonic and one wave type ($m = 0$ and $q = 0$) in the resonator. The electromagnetic field in the continuous metal part of the waveguide $z > h$ was matched to the field in the adjacent semiconductor material. Below, we omit the indices m and q of the y component of the magnetic field. The above approximation is valid when the wavelength in the corrugated waveguide is longer than the corrugation period [8]:
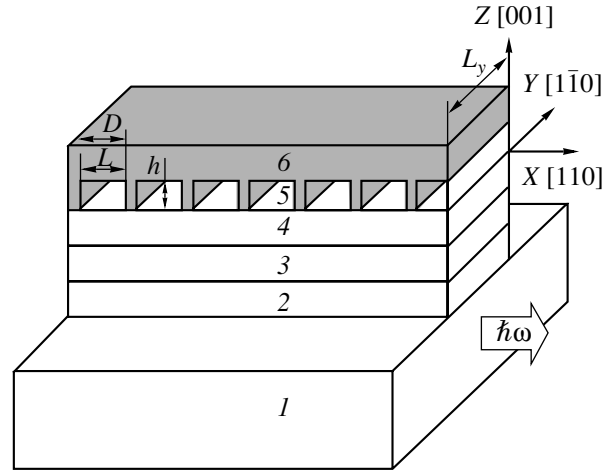
$$2\pi Dn/\lambda < 0.6\pi, \quad (3)$$

where n is the refractive index of the medium and λ is the wavelength in free space, because, under this condition, the amplitude of the zeroth harmonic is significantly larger than those of other harmonics.

The magnetic field intensity $H_y(z)$ of the generated wave in the region $z < 0$ is determined from the following equation:

$$\varepsilon(z, \omega)\frac{d}{dz}\left[\frac{1}{\varepsilon(z, \omega)}\frac{dH_y}{dz}\right] + \left(\varepsilon(z, \omega)\frac{\omega^2}{c^2} - k_x^2\right)H_y$$
$$= -2\varepsilon^{(2)}\frac{k_x\omega}{c}A_1^*(z)A_2(z). \quad (4)$$

Here, the z axis is parallel to the [001] crystallographic direction (Fig. 1) and $\varepsilon^{(2)}$ is the nonlinear susceptibility. To find $H_y(z)$ in the region $z > 0$, we take k_x in Eq. (4) equal to the corresponding projection of the wave vector of the standing waves in the semiconductor parts of the periodic structure, k_{x0}, which is calculated from the relation $k_{x0}^2 + k_{z0}^2 = \varepsilon_s\omega^2/c^2$, where ε_s is the permittivity of the semiconductor and the projection k_{z0} is determined as an eigenvalue of the equation

$$\varepsilon(x, \omega)\frac{d}{dx}\left[\frac{1}{\varepsilon(x, \omega)}\frac{df_q}{dx}\right] + \left(\varepsilon(x, \omega)\frac{\omega^2}{c^2} - k_{zq}^2\right)f_q = 0 \quad (5)$$

for $0 < z < h$ under the boundary condition $f_q(x) = \exp(ik_xD)f_q(x + D)$.

The boundary conditions at $z = 0$ are the continuity of $H_y(z)$ and of

$$\frac{1}{\varepsilon(x, \omega)}\frac{\partial H_y(z)}{dz}.$$

The dependences of the electric field amplitudes A_1 and A_2 of the higher-frequency modes on coordinates and the difference $k_x = k_2 - k_1$ of the modes' propagation
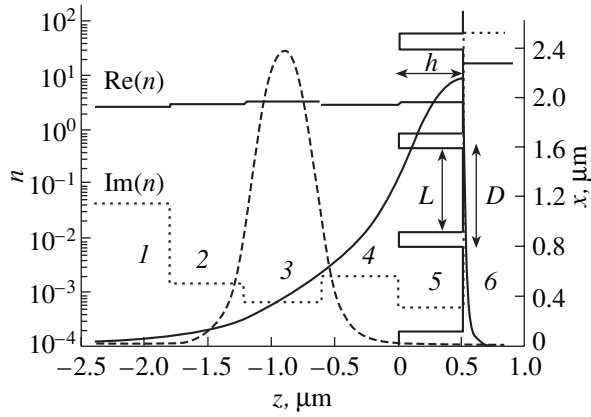
Fig. 2. Profiles of the magnitude of the magnetic field of the difference mode (the dashed line is the amplitude of the high-frequency modes in arbitrary units), the corrugated surface (right), and the real and imaginary parts of the refractive index at a wavelength of 12.7 μm (layer numbers correspond to those used in table).

constants are calculated by solving the wave equation with the corresponding profile of the refractive index.

The electric field component E_z of the difference mode and its power are determined from the equalities

$$E_z = -\frac{1}{\varepsilon(z,\omega)}\left(\frac{ck_x}{\omega}H_y + 2\varepsilon^{(2)}A_1^*(z)A_2(z)\right), \quad (6)$$

$$P = -\frac{cL_y}{2\pi}\int_{-\infty}^{0}\mathrm{Re}(H_yE_z^*)dz, \quad (7)$$

where L_y is the width of the contact strip of the laser diode.

The solution to the above system is a slow wave, which exponentially decays with distance away from the corrugated surface in proportion to the slowing factor (Fig. 2).

Heterostructure parameters

Layer no.	Material	Layer thickness, μm	Doping		Mobility, cm²/V s
			type	concentration, cm⁻³	
1	GaAs	–	n	2×10^{18}	3024
2	InGaP	0.6	n	4×10^{16}	1947
3	GaAs, 2 KYa	0.6	n	4×10^{16}	5796
4	InGaP	0.6	p	1×10^{17}	40
5	GaAs	0.5	p	1×10^{17}	279
6	Au	1	–	–	–

The parameters of the heterostructure proposed here to produce the difference frequency in the InGaAs/GaAs/InGaP system are listed in the table. To produce short-wave modes at two different frequencies, the active region contains two InGaAs quantum wells (QWs) of different depth in the GaAs layer (layer 3 in table). For the short-wave radiation, the waveguide is formed by sandwiching a layer of narrow-gap GaAs between emitter layers of wide-gap InGaP with a lower refractive index. In our calculations, we used the following expressions for the refractive index of the short-wave modes [9, 10]:

$$n_{\mathrm{CaAs}}(\hbar\omega) = \sqrt{7.1 + \frac{3.78}{1 - 0.18(\hbar\omega)^2}}, \quad (8)$$

$$n_{\mathrm{InGaP}}(\hbar\omega) = n_{\mathrm{CaAs}}(\hbar\omega) - 0.279.$$

The effective refractive index of the difference-polarization wave $n = (k_{2x} - k_{1x})/(\omega_2 - \omega_1)$ is determined by the design of the high-frequency part of the laser and is actually independent of the design of the corrugated waveguide, because the high-frequency modes almost do not penetrate beyond the bounding InGaP layers (Fig. 2).

To calculate correctly the refractive index at the difference frequency, the contributions of free carriers and optical photons to the permittivity must be taken into account. The simplest way to consider the contribution of the free-carrier plasma and optical photons to the permittivity is to use expression (26) from [11], which is applicable to both *n*-type and *p*-type semiconductors:

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{\omega_{\mathrm{TO}}^2(\varepsilon_0 - \varepsilon_\infty)}{\omega_{\mathrm{TO}}^2 - \omega^2 - i\Gamma\omega} - \frac{\omega_p^2\varepsilon_\infty}{\omega^2 + i\gamma\omega}, \quad (9)$$

where ε_0 and ε_∞ are the low-frequency and high-frequency permittivities of an undoped semiconductor, respectively; ω_{TO} is the frequency of the transverse optical photon; Γ is the coefficient of the phonon-induced wave attenuation; $\gamma = q/m^*\mu$; $\omega_p^2 = 2\pi nq^2/m^*\varepsilon_\infty$ is the squared plasma frequency; n and m^* are the concentration and effective mass of charge carriers, respectively; and μ is the carrier mobility.

The values of Γ and ω_{TO} were taken from review [11], and γ was calculated from the dependence of the mobility on dopant concentration reported in [9]. The frequency dependence of the attenuation coefficient obtained from (9) is in good agreement with the experimentally observed dependencies [11], except for the multiphonon effects, whose contribution to the attenuation is small.

Our calculations of the permittivity of InGaP in the mid- and far-IR ranges assumed that half of the TO phonons in the solid solution are due to the InP sublattice and half are due to the GaP sublattice, i.e., we neglected the change in the interatomic distances in the solid solution as compared with the binary semicon-
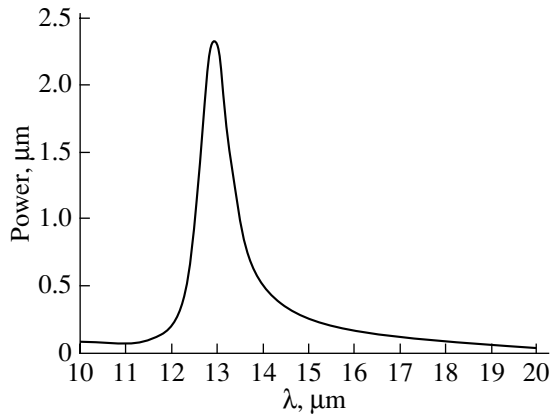
Fig. 3. Power of the difference mode vs. its wavelength in the structure with the following parameters: $h = 0.5$ μm, $D = 0.8$ μm, $L = 0.7$ μm, $\varepsilon^{(2)} = 1.7 \times 10^{-8}$ cm/V, and $L_y = 100$ μm.
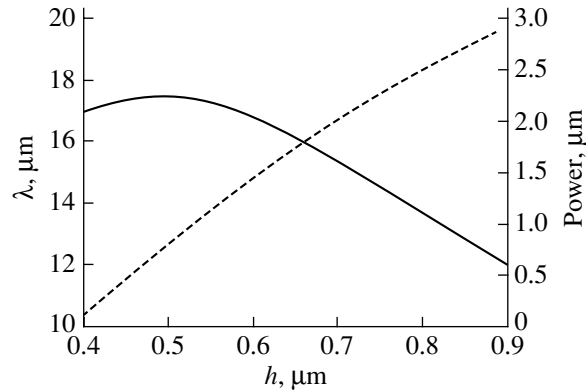

Fig. 4. Maximum power (solid line) and wavelength (dashed line) of the difference mode vs. corrugation depth h at all other parameters fixed (in accordance with table).

ductors InP and GaP. The total permittivity was calculated as a half-sum of the permittivities of InP and GaP. It should be noted that, to reduce loss in the structure, we used a lighter doping than usual.

The power calculated as a function of the difference-mode wavelength is shown in Fig. 3. In these calculations, the maximum wavelength of the high-frequency modes was fixed and the difference-mode wavelength was changed by changing the wavelength of the other high-frequency mode. The function has a maximum, which corresponds to the phase matching between the waveguide and polarization modes at the difference frequency. In addition, it can be seen from Fig. 4 that the power peak shifts to longer wavelengths with increasing the corrugation depth and reaches a maximum value at corrugation depth $h \sim 0.5$ μm.

Although the nonlinear polarization is produced using only fundamental modes, the power of the propagating wave appears to be lower than in [6] by an order of magnitude. The reason is that the coefficient of absorption by free carriers in the metal is significant, since the field amplified exponentially decays with distance away from the corrugated metal surface and is mostly concentrated in the metal resonators of the corrugated waveguide, where they form a standing wave (Fig. 2). The energy of the standing wave exceeds the energy of the propagating wave by a factor of almost five.

Due to the high absorption coefficient, the dependence of the power of the wave in the corrugated waveguide versus the energy of its quantum has a rather wide maximum. In particular, in the vicinity of ≈10 meV, the estimated wave power drops by only an order of magnitude. This means that the generated power is not very sensitive to the parameters of the system in a rather wide frequency range.

Thus, the injection laser of the design proposed in this paper makes it possible to obtain lasing power of

about several microwatts in the mid-IR range at room temperature. The nonlinear polarization is produced using only fundamental short-wave modes in the range of 1 μm. The phase matching is obtained by slowing the long-wavelength IR radiation in the corrugated waveguide.

One of the advantages of the design proposed is its low sensitivity to the spread of the process parameters, because phase matching is obtained by properly choosing the corrugation depth for the difference frequency, which may be determined before applying a corrugation waveguide to the heterostructure. In addition, the optimal corrugation depth may be found empirically using one heterostructure sample.

ACKNOWLEDGMENTS

REFERENCES

1. J. Faist, D. Hofstetter, M. Beck, *et al.*, IEEE J. Quantum Electron. **38**, 533 (2002).

2. M. Rochat, L. Ajili, H. Willenberg, *et al.*, Appl. Phys. Lett. **81**, 1381 (2002).

3. A. Andronov and E. Gornik, Opt. Quantum Electron. **23**, S111 (1991).

4. I. V. Altukhov, M. S. Kagan, K. A. Korolev, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 89 (1999) [JETP **88**, 51 (1999)].

5. A. A. Belyanin, F. Capasso, V. V. Kocharovsky, *et al.*, Phys. Rev. A **63**, 53803 (2001).

6. V. Ya. Aleshkin, A. A. Afonenko, and N. B. Zvonkov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **35**, 1256 (2001) [Semiconductors **35**, 1203 (2001)].

7. A. A. Afonenko, V. Ya. Aleshkin, and A. A. Dubinov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **38**, 244 (2004) [Semiconductors **38**, 239 (2004)].

8. R. A. Silin, *Periodic Waveguides* (Fazis, Moscow, 2002) [in Russian].

9. A. Dargys and J. Kundrotas, *Handbook on Physical Properties of Ge, Si, GaAs and InP* (Science and Encyclopedia, Vilnius, 1994).

10. M. Ohkubo, T. Ijichi, A. Iketani, *et al.*, IEEE J. Quantum Electron. **30**, 408 (1994).

11. J. S. Blakemore, J. Appl. Phys. **53**, R123 (1982).

Translated by A. Khzmalyan

OPTICS,
QUANTUM ELECTRONICS

Analytical Model of a Plasma Photoconverter

N. A. Gorbunov* and G. Flamant**

* Fok Research Institute of Physics, St. Petersburg State University,
Ul'yanovskaya ul. 1, Petrodvorets, St. Petersburg, 198904 Russia
** Institut de Science et de Génie de Matériaux at Procédés, BP5-Odeilo-F66125,
Font-Romeu, Cedex-France
e-mail: gorbunov@paloma.spbu.ru
Received March 4, 2004

Abstract—The problem of using a photoplasma to directly convert the energy of light into electric energy is investigated theoretically. It is shown that the photo-emf is generated due to the ambipolar potential difference caused by the nonuniform ionization of alkali metal vapor by optical radiation. The current–voltage characteristic is calculated and the efficiency is estimated of a plasma photoconverter with plane electrodes. © *2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The most efficient method for reducing the cost of photoelectric energy is to use intense optical radiation [1]. Besides the use of semiconductor photoelectric converters, new technologies for the direct conversion of intense optical radiation into electric current are now being actively investigated. A scheme for converting intense CO_2-laser radiation (with an intensity of a few kW/cm^2) into electric energy by using a thermoemission converter was developed in [2]. An alkali metal thermal-to-electric converter (AMTEC) facility is proposed to convert solar radiation [3]. The AMTEC operation is based on the separation of charges by a membrane that transmits sodium ions well but does not transmit electrons.

The use of a low-temperature plasma for direct optical-to-electric energy conversion has not yet been adequately studied. In [4], a method of magnetohydrodynamic (MHD) conversion of solar energy into electric current by using isothermal alkali metal plasma was proposed. Another method for the direct conversion of the kinetic energy of charged plasma particles into electric current is to use the ambipolar potential difference. This method was discussed in connection with the problem of enhancing the fuel efficiency in nuclear fusion reactors [5].

The first experimental studies on the use of the ambipolar potential difference in an alkali metal photoplasma to directly convert intense optical radiation into electric current were performed in [6, 7]. Plasma formation in alkali metal vapor is a well-studied phenomenon [8]. The efficiency of this process increases by several orders of magnitude when the wavelength of the incident radiation corresponds to the absorption line of the alkali atoms. In the early experiments on photoplasmas, it was shown that the efficiency of optical ionization can approach unity [9]. An interesting result of

[6, 7] was that, in those experiments, the photo-emf was as high as 3–4 V, which was almost one order of magnitude higher than that for semiconductor converters. Such a high photo-emf is a consequence of the high electron temperature in photoplasma [10].

In the present paper, a theoretical model is developed that describes the photo-emf effect in plasma under the action of intense optical radiation. We consider a configuration with plane electrodes, which is the simplest from the standpoint of describing the ambipolar field in plasma.

CALCULATION OF THE CURRENT–VOLTAGE CHARACTERISTIC OF THE CONVERTER

We consider the following simplified mathematical model of the photo-emf effect in plasma under the action of intense optical radiation: Let the plasma be located between two plane electrodes A and B separated by a distance L. We assume that optical excitation occurs in a narrow layer, within which the excited atoms are mainly concentrated. The excitation region is located at a distance x_0 from one of the electrodes. In our analysis, we will ignore the edge effects and will consider the problem in a one-dimensional approximation. It is well known that the dominant ionization mechanism in photoplasma is the electron-impact ionization of resonantly excited atoms [8]. Hence, the ionization source can be represented in the form of a delta function $Q\delta(x - x_0)$, where Q is the ionization rate. In the absence of volume recombination, the diffusion equation for the plasma density n has the form

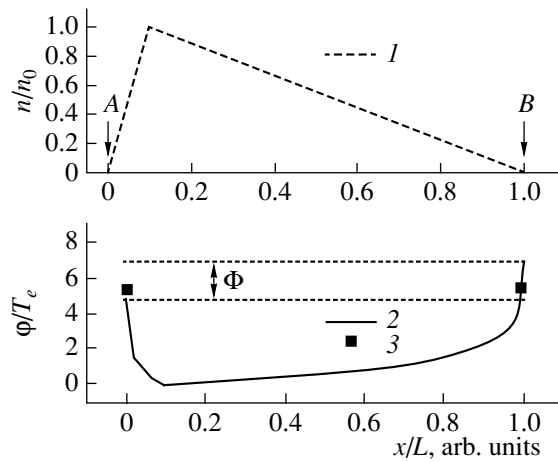$$D_a \frac{d^2 n}{dx^2} = Q\delta(x - x_0). \qquad (1)$$

Fig. 1. Spatial distributions of the (*1*) plasma density and (*2*) electric potential in the open-circuit regime and (*3*) the electrode potential drops in the short-circuit regime at $x_0/L = 0.1$, $\lambda_i/L = 0.02$, $T_i/T_e = 0.1$.

Here, D_a is the coefficient of ambipolar diffusion. In the case of equilibrium electron and ion energy distributions, we have $D_a = D_i(1 + T_e/T_i)$, where D_i is the ion diffusion coefficient and T_e and T_i are the electron and ion temperatures, respectively. The solution to Eq. (1) with zero boundary conditions $n(0) = n(L) = 0$ has the form (see Fig. 1, curve *1*)

$$n(x) = \frac{n_0}{x_0}x \quad \text{for} \quad 0 \le x \le x_0, \tag{2a}$$

$$n(x) = \frac{n_0}{L - x_0}(L - x) \quad \text{for} \quad x_0 \le x \le L. \tag{2b}$$

The plasma density at the point x_0 is equal to $n_0 = Q(L - x_0)x_0/D_aL$. Equation (1) becomes invalid at a distance of $x_0 \approx \lambda_i$ from the electrode surface, where λ_i is the ion mean free path. The potential drops φ_A and φ_B across the electrode sheaths partially cut off the electron diffusive fluxes toward the electrodes. Kirchhoff's laws for the electric circuit with unit-area electrodes have the form

$$I(\varepsilon) = i_A - j_A = -i_B + j_B, \\ \varepsilon(I) = \phi_A + \varphi_A - \phi_B - \varphi_B. \tag{3}$$

Here, $I(\varepsilon)$ is the current in the external circuit; ε is the interelectrode voltage; j_A and j_B are the electron current densities at electrodes A and B, respectively; and i_A and i_B are the current densities of positive ions. If the circuit closing electrodes A and B consists only of a resistance R, then we have $I = \varepsilon/R$. The ambipolar potential drop ϕ across the quasineutral plasma is determined by the expression [11]

$$\phi = T_e \int \frac{1}{n}\frac{dn}{dx}dx. \tag{4}$$

Hence, we find that $\phi_A = T_e\ln(x_0/\lambda_i)$ for $\lambda_i \le x \le x_0$ and $\phi_B = T_e\ln((L - x_0)/\lambda_i)$ for $x_0 \le x \le L - \lambda_i$.

The ion current density is determined by the ambipolar flow velocity $i = -D_a dn/dx$. Hence, we find that $i_A = -D_an_0/x_0$ and $i_B = D_an_0/(L - x_0)$. The electron current density at the electrode can be find under the assumption that the electrode sheath is collisionless, since the electron mean free path usually far exceeds λ_i. When the electron energy distribution is Maxwellian, the electron current density is expressed through the well-known Langmuir formula

$$j(e\varphi) = \frac{n_b e \overline{v}}{4}\exp\left(-\frac{e\varphi}{kT_e}\right), \tag{5}$$

where n_b is the plasma at the sheath boundary and $\overline{v} = (8kT_e/\pi m)^{1/2}$ is the thermal electron velocity.

The solution to set of equations (3) yields the current–voltage characteristic of the photo-emf source:

$$\varepsilon = T_e \ln\left[\frac{\dfrac{D_an_0}{x_0} + I}{\dfrac{D_an_0}{L - x_0} - I}\right]. \tag{6}$$

Let us examine the limiting cases of expression (6). In the open-circuit regime ($I(\varepsilon) = 0$), the electrode potential drops are equal to one another, $\varphi_A^0 = \varphi_B^0 \approx T_e\ln\sqrt{MT_i/mT_e}$, and the electron and ion fluxes to the electrodes are also equal. In this case, the photo-emf Φ is determined by the ambipolar potential difference

$$\Phi = T_e \ln\left(\frac{L - x_0}{x_0}\right). \tag{7}$$

It follows from formula (7) that the emf vanishes when the electrodes are positioned symmetrically about the ionization region ($x_0 = L/2$). Curve *2* in Fig. 1 shows the potential distribution and the electrode potential drops in the open-circuit regime for a lithium vapor plasma.

When the current flows through the external circuit, the potential drop near the electrode that is closest to the ionization source increases in comparison to φ_A^0 and a greater ion flux arrives at this electrode from the plasma. An additional electron flux arrives at this electrode through the external circuit. The potential of the electrode located farther from the ionization source decreases, and more electrons than ions arrive at this electrode from the plasma. Thus, the electrode potential drops play the role of a peculiar kind of valves controlling the electron fluxes to the electrode. The ambipolar potential difference in our model remains unchanged.

In the short-circuit regime ($\varepsilon = 0$), the current I_0 is determined by the expression

$$I_0 = Q\frac{L - 2x_0}{2L}. \qquad (8)$$

The electrode potential props in this case are shown in Fig. 1 by squares *3*.

Figure 2 shows the calculated current–voltage characteristic $\varepsilon(I)$ and the power $W = \varepsilon I$ released in the external circuit for two different positions of the ionization regions with respect to the electrodes. It can be seen from the figure that the photo-emf in the open-circuit regime, the short-circuit current, and the power released in the load increase as the ionization region approaches one of the electrodes. Note that, in Fig. 2, the photo-emf is normalized to T_e, and the current is normalized to the ionization rate in the plasma. To determine T_e, it is necessary to consider equations describing the plasma state.

Fig. 2. (*1*, *2*) The emf and (*3*, *4*) output power of a photoconverter as functions of the electric current at different positions of the ionization source: $x_0/L = $ (*1*, *3*) 0.1 and (*2*, *4*) 0.01.

MODEL OF A PHOTOPLASMA IN THE DIFFUSION REGIME

Let us estimate T_e from equations for photoplasma. The relative population of the resonantly excited states is characterized by the effective temperature

$$T^* = -\Delta E_{01}[\ln(N_1 g_0/N_0 g^*)]^{-1},$$

where ΔE_{01} ($1.53 \leq \Delta E_{01} \leq 2.1$ eV) is the energy difference between the resonant levels of alkali atoms, N_1 and N_0 are the populations of alkali atoms in the excited and ground states, and g_0 and g^* are the statistical weights of the corresponding levels.

The main source of electron heating in photoplasma is superelastic collisions with resonantly excited atoms. Far from the optical saturation of the resonance transition ($T^* \leq \Delta E_{01}$), the frequency of electron-impact excitation from the ground state $\nu_{01}(T_e)$ is much higher than the frequency of ionization from the resonance state, $\nu_{1i}(T_e) \ll \nu_{01}(T_e)$. This is a direct consequence of the relations between the cross sections and the energy thresholds for the above processes: $\Delta E_{01} < \Delta E_{1i}$, where ΔE_{1i} is the energy of ionization from the ground state. As a result, the electron energy fraction spent on the ionization of atoms is less than 1% of the energy spent on the excitation of atoms from the ground state. The electron energy lost in inelastic processes (at $T_e >$ 0.15 eV) is larger than the energy lost in elastic electron–atom collisions:
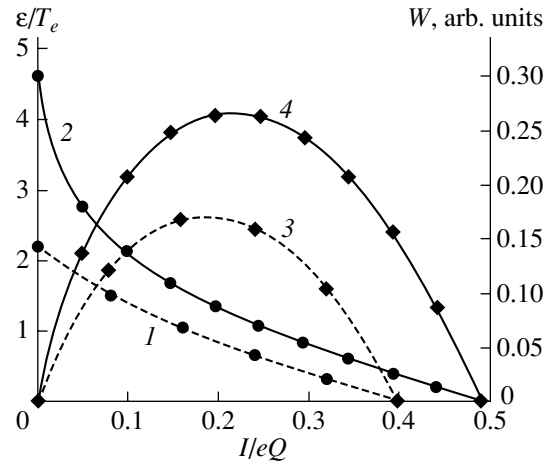
$$\Delta E_{01}\nu_{01}(T_e) > \delta\nu_{ea}(T_e)(T_e - T_g),$$

where $\delta = 2m/M$ is the coefficient of energy transfer in elastic collisions of electrons with atoms, $\nu_{ea}(T_e)$ is the frequency of elastic collisions, and T_g is the gas temperature [12].

The governing role of the processes of the resonant excitation and deexcitation in the electron energy balance equation is the reason why $T_e \approx T^*$. This is confirmed by numerical calculations of the energy distribution function in photoplasma [13].

The electron density n_e and the population of alkali atoms in the excited state N_1 are determined from the corresponding balance equations. In evaluating the maximum possible conversion efficiency, we ignore the recombination of electrons and ions in the plasma in comparison to their diffusive losses. Because of the trapping of the resonance radiation, the effective lifetime of the resonance level substantially exceeds its natural lifetime. Let us consider the conditions under which the rate of the electron-impact deexcitation of the resonantly excited states exceeds the radiative decay rate. Under the above assumptions, the balance equations take the form

$$\frac{\partial n_e}{\partial t} = k_{1i}(T_e)n_e N_1 - \frac{n_e}{\tau_{\mathrm{ad}}(T_e)},$$

$$\frac{\partial N_1}{\partial t} = \frac{J}{\Delta E_{01}} + k_{01}(T_e)n_e N_0 - k_{10}(T_e)n_e N_1 - k_{1i}(T_e)n_e N_1, \qquad (9)$$
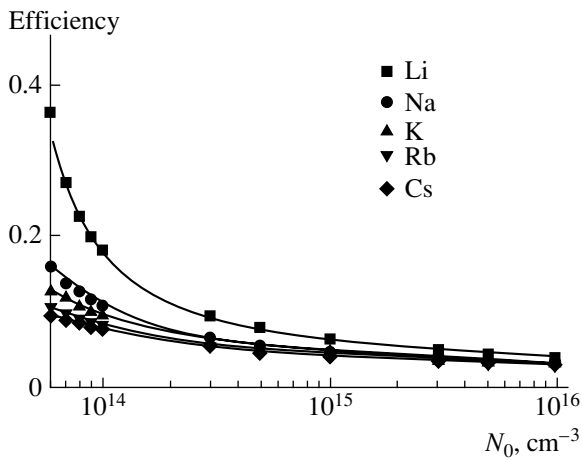
Fig. 3. Conversion efficiency for different alkali metals as a function of the vapor density for $L = 1$ cm and $x_0/L = 0.1$.

where J [J/(cm^3 s)] is the rate of the external optical excitation of the resonant levels (in units of J/(cm^3 s)) and $\tau_{ad}(T_e) = x_0(L - x_0)/D_a$ is the characteristic time of ambipolar-diffusion.

The first of the above equations allows us to determine T_e, assuming that $T_e = T^*$. Here, as in the Shottky theory, the electron temperature is determined by the ionization balance condition. The difference from conditions in an electric discharge is that the ionization from the ground atomic state is replaced with the ionization from the resonantly excited state. Using the Seaton formula [14] for $k_{1i}(T_e)$ and expressing D_i in terms of the resonance charge-exchange cross section σ_{res} [15] and T_g, we derive the following equation for determining T_e:

$$T_e = \frac{E_{0i}}{\ln\!\left(\alpha \dfrac{\sqrt{T_e T_g M}}{E_{1i}^2} \dfrac{N_0^2 \sigma_{res}}{T_g + T_e} x_0(L - x_0)\right)}, \quad (10)$$

where α is a factor depending on the system of units used in formula (10).

It follows from this expression that, for the same external parameters N_0, x_0, and L, the T_e value is higher in the photoplasma of light alkali metals, because these metals have the higher ionization energies E_0 and E_{1i} and the lower values of M and σ_{res}. The solution to Eq. (10) has a singularity at $x_0 \sim \lambda_i$. At $x_0 < \lambda_i$, Eq. (10) has no solutions, which is a consequence of the zero boundary conditions used in solving Eq. (1).

The second of Eqs. (9) allows us to determine n_0. Assuming that $T_e = T^*$, we obtain

$$n_0 = \frac{J\tau_{ad}(T_e)}{\Delta E_{01}}. \quad (11)$$

It follows from this formula that the degree of ionization increases as the radiative-excitation power and the distance of the ionization region from the electrode surface increase.

CONVERSION EFFICIENCY FOR THE ABSORBED RADIATION ENERGY

The dependence of the output power on the current through the load (Fig. 2) shows that the optimal current is nearly equal to one-half of the short-circuit current. This current determines the lower estimates for the maximum possible power ($P_{opt} = 0.5 I_{max}\varepsilon(0.5 I_{max})$) and the conversion efficiency η:

$$\eta = \frac{P_{opt}}{J} = \frac{1}{4}\frac{L - 2x_0}{L}\frac{T_e}{\Delta E_{01}}\ln\!\left(\frac{3L - 2x_0}{L + 2x_0}\right). \quad (12)$$

The calculated conversion efficiencies for the plasmas of alkali metals as functions of their vapor densities are shown in Fig. 3. When calculating T_e from Eq. (10), it was assumed that the density of alkali atoms $N_0(T_g)$ corresponded to the saturation vapor pressure [16]. It can seen that, under identical external conditions, the maximum efficiency is achieved in lithium vapor because of the maximum values of T_e in it.

The maximum possible efficiency of a photoconverter with plane electrodes ($x_0 \ll L$) can be estimated by the formula

$$\eta = \frac{P_{opt}}{J} = \frac{1}{4}\frac{T_e}{\Delta E_{01}}\ln 3. \quad (13)$$

It follows from this formula that, to increase the conversion efficiency, the electron temperature should be fairly high.

CONCLUSIONS

In this paper, an analytical model of a photoconverter with plane electrodes has been developed. It is shown that the photo-emf effect in plasma is caused by the generation of an ambipolar field when the electrodes are positioned asymmetrically about the region of gas excitation by an external radiation source. The electric current in the external circuit is determined by the diffusion of the positive ions (i.e., as in semiconductors, by the diffusion of the minority current carriers). An increase in the electron temperature leads to an increase in both the photo-emf and the conversion efficiency.

REFERENCES

1. V. M. Andreev, V. A. Grilikhes, and V. D. Rumyantsev, *Photoelectric Conversion of Concentrated Solar Radiation* (Nauka, Leningrad, 1989).

2. A. P. Budnik and V. A. Zherebtsov, Zh. Tekh. Fiz. **71** (6), 60 (2001) [Tech. Phys. **46**, 704 (2001)].

3. J. Terry, K. Hendricks, and C. Huang, J. Sol. Energy Eng. **122** (2), 49 (2000).

4. G. J. Dunning and A. J. Palmer, J. Appl. Phys. **52**, 7086 (1981).

5. I. Mori and K. Sumitomo, IEEE Trans. Plasma Sci. **16**, 623 (1988).

6. N. A. Gorbunov and T. Stacewicz, Pis'ma Zh. Tekh. Fiz. **26** (15), 21 (2000) [Tech. Phys. Lett. **26**, 654 (2000)].

7. N. A. Gorbunov and T. Stasevich, Teplofiz. Vys. Temp. **39**, 668 (2001).

8. I. M. Beterov, A. V. Eletskiĭ, and B. M. Smirnov, Usp. Fiz. Nauk **155**, 265 (1988) [Sov. Phys. Usp. **31**, 535 (1988)].

9. V. B. Brodskiĭ and A. G. Voronchev, Zh. Tekh. Fiz. **40**, 1927 (1970) [Sov. Phys. Tech. Phys. **15**, 1501 (1970)].

10. N. A. Gorbunov, A. Grochola, P. Kruk, *et al.*, Plasma Sources Sci. Technol. **11**, 492 (2002).

11. V. A. Rozhanskiĭ and L. D. Tsendin, *Collisional Transport in Partially Ionized Plasma* (Énergoatomizdat, Moscow, 1988).

12. N. Ya. Shaparev, Zh. Tekh. Fiz. **49**, 2229 (1979) [Sov. Phys. Tech. Phys. **24**, 1230 (1979)].

13. N. A. Gorbunov, F. E. Latyshev, and A. S. Mel'nikov, Fiz. Plazmy **24**, 950 (1998) [Plasma Phys. Rep. **24**, 885 (1998)].

14. L. A. Vaĭnshteĭn, I. I. Sobelman, and E. A. Yukov, *Excitation of Atoms and Broadening of Spectral Lines* (Nauka, Moscow, 1979; Springer-Verlag, Berlin, 1981).

15. B. M. Smirnov, Usp. Fiz. Nauk **171**, 233 (2001) [Phys. Usp. **44**, 221 (2001)].

16. *Handbook of Physical Quantities*, Ed. by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991).

Translated by N. Larionova

============================ **RADIOPHYSICS** ============================

Conducting and Reflecting Properties of Thin Metal Films

I. V. Antonets, L. N. Kotov, S. V. Nekipelov, and E. N. Karpushov

Syktyvkar State University, Syktyvkar, 167001 Russia
e-mail: kotov@syktsu.ru
Received March 16, 2004

Abstract—Amorphous silver, copper, gold, and iron films with a thickness of 6–60 nm have been grown on a polymer substrate by the method of vacuum deposition. The dependences of the specific conductivity and the microwave reflection coefficient on the film thickness are obtained and a relation between these values is established. © *2004 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

A great number of papers are devoted to investigations of the conductivity and electrodynamic properties of thin (10- to 100-nm-thick) metal films, in particular, to the reflection, transmission, and absorption of electromagnetic waves in such films [1, 2]. These investigations have been mostly carried out under the assumption that the size of inhomogeneities in the films is much smaller than the wavelength but much greater than the skin depth, which implies that the wave field distribution inside the film is inhomogeneous. For a film thickness of 10–100 nm, this assumption is valid only in the range of rather high frequencies in the visible and infrared ranges [1–13]. Less attention has been paid to the electrodynamic properties in the centimetric range, for which the skin depth is much greater than the film thickness. However, this very range is of interest from the standpoint of using metal films in various data processing devices, where the metal films are promising candidates to replace expensive ferrite films [14, 15].

The present work was devoted to investigation of the conductivity and the reflection of centimetric electromagnetic waves as functions of the thickness of a thin metal film of silver, copper, gold, and iron in the amorphous state. We have also established a relationship between these functions.

1. CHARACTERIZATION OF SAMPLES

We have studied the amprophous films of silver, copper, gold, and iron obtained by thermal deposition in vacuum onto the surface of a dielectric polymer substrate at room temperature. The vacuum was about 10^{-5} Torr and the metal deposition rate was ~10 nm/s. Such a high rate was necessary to obtain pure metal films with a small content of impurities in a rough vacuum. These deposition rates are often used to fabricate thin-film materials in industry. The film thickness was varied from 6 to 60 nm.

The X-ray diffraction analysis of the films showed that diffraction peaks typical of the crystalline state were entirely absent; hence, the films were X-ray amorphous. A polymeric X-ray film with a thickness of 0.5 mm was used as a substrate for depositing metal films used in the measurements of conductivity and reflection coefficient.

2. EXPERIMENTAL EQUIPMENT AND TECHNIQUES

2.1. Determination of the film thickness and conductivity. A crystalline calibrator assembled on the base of a quartz resonator was used to control film thickness during the deposition. The evaporating material was simultaneously deposited on the substrate and facets of the quartz crystal linked to the high-frequency oscillatory circuit of a generator. The procedure of film thickness measurements was described in [16].

The conductivity measurements were performed using an experimental setup schematically depicted in Fig. 1, which consisted of dc power supply B5-43, digital wavemeter Ch3-57, digital ohmmeter Shch-34, and piezoquartz transducer (calibrator) placed in the vacuum chamber at the same level as the substrate. Thick silver layers were deposited at the ends of each polymer substrate, which were connected to the measuring equipment. The resistance was measured after each deposition cycle.

During the deposition, all the experimental data (deposition time, piezoquartz transducer frequency, and resistance) were recorded into computer memory (writing speed was about 200–250 experimental points per minute). In some experiments, a thin metallic grid was placed between the evaporating material and substrate so as to decrease the deposition rate, thus increasing the amount of the measured data. The film thickness was determined from the frequency shift of the quartz resonator, while, taking into account the thickness, area, and resistance, it was possible to determine the conductivity.

2.2. Determination of the reflection coefficient.

The coefficient of reflection of a microwave signal from metal films was measured at normal incidence. The measurements were carried out in a frequency range of 8–12 GHz. The sample film was placed in a rectangular waveguide with a cross-section of 10×24 mm and oriented perpendicular to its longitudinal axis so as to cover all the waveguide cross-section. A matched load was placed behind the film, at the opposite end of the waveguide (Fig. 2). The waveguide was connected to a panoramic device measuring the complex transfer constants. The device comprised a swept-frequency generator (SFG-61), a voltage standing-wave indicator with an attenuation module (YaSR-67), and a waveguide set of reflectometers. In order to increase the sensitivity, the microwave signal was modulated in amplitude with a frequency of 100 kHz. The initial signal from the microwave generator was divided and fed into two channels. One of these signals was used as a reference signal and entered the indicator immediately after being detected, while the other signal was fed to the waveguide containing the investigated sample. The reflected wave excited an electromotive force (emf) in the receiver. The response signal was detected and fed to the indicator. In both cases, detection was carried out in such a way that the output signal was proportional to the microwave field strength. Owing to this connection, it was possible to compare the amplitudes of the wave field strength of the incident and reflected signals in the indicator. The reflection coefficient R was determined as the ratio of these signals,

$$R = \frac{K-1}{K+1} = \frac{E_{\text{refl}}}{E_{\text{inc}}},$$

where K is the standing-wave ratio (determined directly on the indicator scale) and E_{inc} and E_{refl} are the amplitudes of the incident and reflected waves, respectively.

3. MAIN EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Electric Conductivity

The experimental data on the specific conductivity σ as a function of the film thickness are shown in Fig. 3 for various metal films measured at room temperature in a vacuum of 10^{-5} Torr (for iron films, the values of specific conductivity are multiplied by a factor of five). These experimental data are well approximated by the Boltzmann function (Fig. 3). As follows from Fig. 3, the dependence of conductivity on the thickness is weak for all films with a thickness of up to 5–7 nm. For the films with such thicknesses, a cluster-island structure is typical [10, 11, 17], which is characterized by a high resistivity along the film. Moreover, objects with such small thicknesses are characterized by a small density of free electrons and a short free path limited by the film surface, clusters, and islands on the film sur-
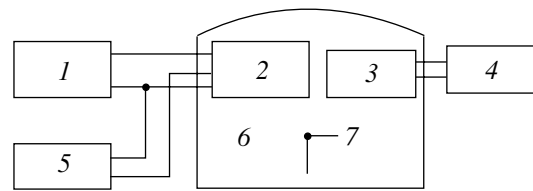
Fig. 1. Schematic diagram of the experimental used for the measurements of the conductivity and thickness of the films: (*1*) power supply; (*2*) piezoquartz transducer; (*3*) substrate; (*4*) ohmmeter; (*5*) wavemeter; (*6*) vacuum chamber; and (*7*) metal source.
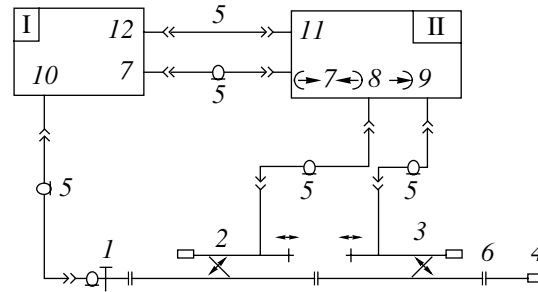

Fig. 2. Schematic diagram of the experimental setup for the measurements of the standing-wave ratio: (I)—swept-frequency generator (SFG); (II) indicator; (*1*) doorknob transformer; (*2*) directional detector of incident wave; (*3*) directional detector of reflected wave; (*4*) matched load; (*5*) interface cable; (*6*) sample; (*7*) ARM outputs of indicator and SFG; (*8*) incident wave jack; (*9*) reflected wave jack; (*10*) SFG output; (*11*) socket connector of SFG; and (*12*) socket connector of indicator.
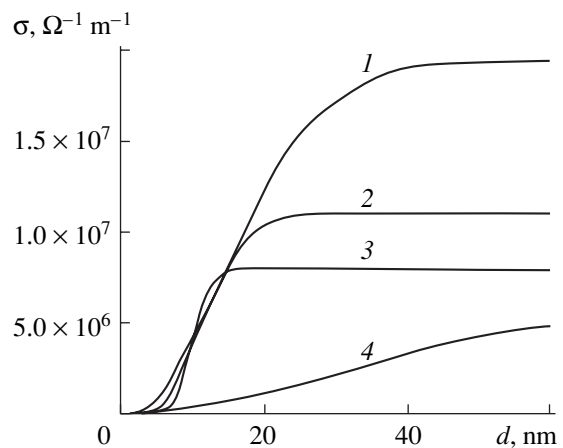

Fig. 3. Plots of the specific conductivity versus film thickness for the films of various metals: (*1*) silver; (*2*) copper; (*3*) gold; and (*4*) iron.

face. All these factors lead to a low value of the longitudinal (lateral) conductivity of the films.

As the film thickness increases, the film conductivity exhibits a sharp dependence on the film thickness (in the range of 7–15 nm for gold films, 6–20 nm for copper films, and 5–40 nm for silver films). For iron films,

this dependence is not as pronounced and the conductivity increases with the film thickness in the range from 5 to 60 nm and above. With further increase in the film thickness, the dependence $\sigma(d)$ becomes less pronounced and the specific conductivity of the films is close to maximum. It should be noted that the character of conductivity growth for silver, copper, and gold films in the indicated range of thicknesses is almost identical, and the conductivity of all these films in the saturation region is close to the corresponding bulk values. In particular, this value is about $2 \times 10^7 \ \Omega^{-1} \ m^{-1}$ for silver films, $1.1 \times 10^7 \ \Omega^{-1} \ m^{-1}$ for copper films, $8 \times 10^6 \ \Omega^{-1} \ m^{-1}$ for gold films, and about $1 \times 10^6 \ \Omega^{-1} \ m^{-1}$ for iron films. The mechanism of conductivity in thin amorphous metal films was recently considered elsewhere [16].

It should be noted that the results of conductivity measurements are different from those recently reported in [16]. This difference is associated with the substrates, on which the metal films were deposited. Unlike a textolite substrate used for the conductivity measurements reported in [16], the polymeric X-ray film had a much smoother surface with smaller surface roughnesses, which was revealed by atomic force microscopy measurements. Hence, the layer thickness on the polymeric X-ray film substrate is greater than that on the textolite surface for the same amount of deposited metal (the values of $\sigma(d)$ for a textolite substrate were reported in [16]).

3.2. Microwave Reflection

3.2.1. Experimental results and discussion. Figure 4 shows the experimental dependences of the reflection coefficient of a 10 GHz microwave signal as a function of the film thickness for silver, copper, gold, and iron films. It is seen from these experimental data that the $R(d)$ curves well correlate with the analogous dependences of the conductivity $\sigma(d)$ (Fig. 3). At a film thickness of about 5–7 nm, the reflection coefficient is
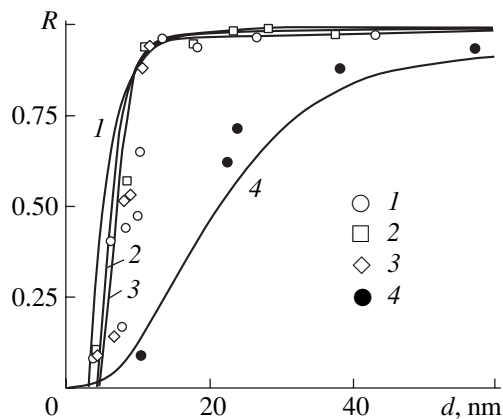

Fig. 4. Theoretical (solid lines) and experimental (points) dependences of the reflection coefficient on the film thickness: (*1*) silver; (*2*) copper; (*3*) gold; and (*4*) iron.

much smaller than unity. This fact indicates that a reflecting layer capable of coherently reflecting microwaves is not formed in very thin amorphous films. Such thin layers are also characterized by low values of the conductivity. As the film thickness increases above 5–7 nm, the reflection coefficient R exhibits a sharp increase for all the films. This is related to the formation of a reflecting layer and to an increase in the film conductivity in this range. Notice that, beginning with a thickness of ~15 nm, the reflectance of silver, copper, and gold films is saturated on a level of $R \longrightarrow 1$. The same film thickness corresponds to saturation of conductivity for gold films (Fig. 3, curve *3*). Hence, a conductivity of about $8 \times 10^6 \ \Omega^{-1} \ m^{-1}$ is sufficient for such a film to almost fully reflect the microwave radiation. Further increase in the conductivity of silver and copper films at a given film thickness (Fig. 3, curves *1*, *2*) does not lead to appreciable changes in the reflection coefficient, which is associated with the development of other effects (for example, absorption) compensating an increase in the conductivity.

For iron films, the reflectance is virtually not saturated with the thickness increasing up to 60 nm. This can be explained by the fact that the conductivity of iron films is $< 8 \times 10^6 \ \Omega^{-1} \ m^{-1}$ even for $d \approx 60$ nm (Fig. 3, curve *4*) and, hence, such films cannot reflect microwaves completely, which is confirmed by Fig. 4. Therefore, for the total reflection, it is necessary that a conductivity of about $10^7 \ \Omega^{-1} \ m^{-1}$ in metal films be reached at very small film thicknesses ($d \approx 10$ nm).

3.2.2. The influence of a medium on the reflecting properties of the films. When a metal film is exposed to air, the reflection coefficient gradually decreases with time. This fact is, apparently, caused by oxidation of the film surface [1, 3, 5]. Thus, the measurement of the microwave reflection coefficient of thin films makes it possible to estimate oxide layer thickness. However, in our case, the time variation of the reflection coefficient can reflect the influence of other factors. In order to exclude this effect, subsequent measurements were carried out for the films kept in air until a stationary state was attained.

For silver and copper films, the fastest drop in the reflection coefficient was observed within a few seconds of exposure to air, after which the drop slowed down and the reflection coefficient reached an almost stationary value. For iron films, this drop could be observed for several hours, that is, the oxidation processes proceeded slower than in other films. For gold films, the exposure to air did not significantly change the reflection coefficient.

As can be seen from Fig. 4, the reflection coefficient at small thicknesses (below 10 nm), for example, for silver films, is lower at some points than that for gold films. This fact can be explained by the fact that the formation of an oxide layer on silver films decreases the effective conductivity and, hence, the reflectance of these films. The conductivity of gold films remains

almost constant. Hence, at these thicknesses, the measured reflection coefficient for gold films can be somewhat higher than that for silver or copper films (Fig. 4).

3.2.3. Comparison of the experimental and theoretical results. Interpretation of the results on microwave reflection from thin films requires using a correct theoretical model capable of associating the measured electrodynamic parameters with material properties. In the investigated frequency range (~10 GHz) the electromagnetic wavelength (~3 cm) is five to seven orders of magnitude greater than the film thickness (6–60 nm). The skin depth in the same range of frequencies at a conductivity of 10^7 Ω^{-1} m^{-1} typical of a bulk metal is 500 nm, which is about one order of magnitude greater than the film thickness. For thin films, especially those in the amorphous state characterized by a lower conductivity, the ratio of the skin depth to the film thickness is even much greater and it can be assumed that the electromagnetic wave field penetrates through the whole film thickness. In such a case, the problem of wave reflection is usually solved by the classical methods [11, 18]. However, the calculations can be significantly simplified by using the averaging method. This method was first applied for calculations of a waveguide with ferrite filling [19], in which the dependence of the wave field on the coordinate perpendicular to the layer was assumed to be linear. This method was further developed for various layers [20–22]; detailed analysis of the applicability of the averaging method in different cases in comparison with the exact solution was given in review [22].

Now, we will apply the averaging method to the problem of wave reflection from a thin metal film at normal incidence. Let the space and time dependence of the incident wave field have the form $\exp[i(\omega t - kz)]$, where z is the coordinate along the normal to a film surface. As follows from [22], the coefficient of wave reflection from a metal film with the thickness d and the conductivity σ can be expressed as

$$R = \frac{\alpha - \beta}{(\alpha + 1)(\beta + 1)}, \qquad (1)$$

where

$$\alpha = \frac{i\omega\mu_0 d}{2\eta_0}, \quad \beta = \frac{\sigma\eta_0 d}{2},$$

μ_0 is the permeability of vacuum and η_0 is the impedance of free space.

For a frequency of 10 GHz, a film thickness of 15 nm, and a conductivity of 10^7 Ω^{-1} m^{-1} we obtain $|\alpha| \sim 10^{-7}$ and $\beta \sim 100$ (i.e., the value of α can be ignored in comparison with β). In this case, to within the sign, we obtain the reflection coefficient in the form

$$R = \left(1 + \frac{2}{\sigma\eta_0 d}\right)^{-1}, \qquad (2)$$

where σ is a function of the layer thickness described by the expression

$$\sigma = \frac{A_1 - A_2}{1 + \exp\left(\dfrac{d - x_0}{dx}\right)} + A_2. \qquad (3)$$

Here, A_1 and A_2 are some material constants, d_0 is the initial point corresponding to the minimum layer thickness, and dx is some interval corresponding to a reference step.

Formula (2) gives the reflection coefficient with respect to the wave field intensity, which is measured in experiments. The reflection coefficient with respect to power is equal to the square of this expression. The reflection coefficient determined by Eq. (2) coincides with that obtained by the classical method [9] (to within the notations and the systems of units), which means that application of the averaging method in the case under consideration is justified.

Figure 4 (solid curve) shows the theoretical curve of $R(d)$ calculated by Eq. (2) with regard to the experimental data for $\sigma(d)$ approximated by Boltzmann dependence (3). It is seen that the curve is a good fit to the experimental data.

Some differences between theoretical and experimental dependences $R(d)$ can be associated with the substrate heterogeneity affecting the surface relief of thin films and with unknown variable oxide layer thickness. For bulk samples, the oxide layer thickness can reach 10 nm. In our experiments, the conductivity was measured in vacuum, while the reflection coefficient was determined in the air. This fact may also cause some changes in the film conductivity during the measurements.

4. CONCLUSIONS

Amorphous films of silver, copper, gold, and iron with a thickness of 6–60 nm were grown by thermal deposition onto a polymer substrate under moderate vacuum at room temperature. The interval of film thicknesses was found in which the specific conductivity exhibits a sharp dependence on the film thickness. The dependences of the microwave reflection coefficient $R(d)$ on the film thickness correlates with the behavior of conductivity $\sigma(d)$. Saturation of the reflectance of silver, copper, and gold films ($R \longrightarrow 1$) was observed at a thickness of about 15 nm, when a reflecting layer was already formed and the film conductivity was high enough for almost total reflection of microwave radiation. For iron films, saturation of the reflectance was not observed in the range of film thicknesses studied. Using the averaging method, we obtained a formula for the reflection coefficient as a function of the film thickness, which well describes the experimental data on the reflection of microwave radiation. If several values of the reflection coefficient in the range of increase and saturation of $R(d)$ are known, the dependence of the

conductivity on the film thickness can be obtained and the type of the film can be determined.

REFERENCES

1. A. V. Sokolov, *Optical Properties of Metals* (Fizmatgiz, Moscow, 1961; Elsevier, New York, 1967).

2. R. F. Soohoo, *Magnetic Thin Films* (Harper and Row, London, 1965; Mir, Moscow, 1967).

3. L. I. Maissel and R. Glang, *Handbook of Thin Film Technology* (McGraw-Hill, New York, 1970; Sov. Radio, Moscow, 1977).

4. D. K. Larson, in *Physics of Thin Films: Advances in Research and Development*, Ed. by G. Haas, M. H. Francombe, and R. W. Hoffman (Academic, New York, 1974; Mir, Moscow, 1977), Vol. 6.

5. F. Abeles, in *Physics of Thin Films: Advances in Research and Development*, Ed. by G. Haas, M. H. Francombe, and R. W. Hoffman (Academic, New York, 1974; Mir, Moscow, 1977), Vol. 6.

6. S. Fahy, C. Kittel, and S. G. Louie, Am. J. Phys. **56**, 989 (1988).

7. P. A. Dowben, Surf. Sci. Rep. **40**, 151 (2000).

8. G. Marchal, P. Mangin, and C. Janot, Thin Solid Films **23**, 17 (1974).

9. A. E. Kaplan, Radiotekh. Élektron. (Moscow) **9**, 1781 (1964).

10. A. K. Sarychev, D. J. Bergman, and Y. Yagil, Phys. Rev. B **51**, 5366 (1995).

11. R. Levy-Nathansohn and D. J. Bergman, Phys. Rev. B **55**, 5425 (1997).

12. M. Fenn, G. Akuetey, and P. E. Donovan, J. Phys.: Condens. Matter **10**, 1707 (1998).

13. M. Aeschlimann, M. Bauer, S. Pawlik, *et al.*, Appl. Phys. A **71**, 485 (2000).

14. J. D. Adam, Proc. IEEE **76**, 159 (1988).

15. W. S. Ishak, Proc. IEEE **76**, 171 (1988).

16. I. V. Antonets, L. N. Kotov, S. V. Nekipelov, and E. A. Golubev, Zh. Tekh. Fiz. **74** (3), 24 (2004) [Tech. Phys. **49**, 306 (2004)].

17. H.-D. Liu, Y.-P. Zhao, G. Ramanath, *et al.*, Thin Solid Films **384**, 151 (2001).

18. L. M. Brekhovskikh, *Waves in Layered Media* (Nauka, Moscow, 1973; Academic, New York, 1980).

19. M. I. Kontorovich and A. S. Cherepanov, Radiotekh. Élektron. (Moscow) **30**, 1543 (1985).

20. M. I. Oksanen, S. A. Tretyakov, and I. V. Lindell, J. Electromagn. Waves Appl. **4**, 613 (1990).

21. M. I. Kontorovich and S. A. Tret'yakov, Radiotekh. Élektron. (Moscow) **31**, 1110 (1986).

22. L. Ya. Khaliulin and S. A. Tret'yakov, Radiotekh. Élektron. (Moscow) **43**, 16 (1998).

Translated by M. Astrov

Influence of Contact Resistance on the Percolation Interval in Disordered Metal–Insulator Macrosystems

V. A. Sotskov

Berbekov Kabardino-Balkarian State University, Nalchik, 360004 Kabardino-Balkaria, Russia
e-mail: sozkov_va@rambler.ru
Received July 23, 2003

Abstract—The dependence of the resistivity of a paraffin–conductor macrosystem on the conducting phase concentration has been experimentally studied. A model of conduction in systems with different contact resistances in terms of the percolation theory is suggested, and an indirect characteristic of contact resistance is introduced. © *2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Present-day percolation theory [1, 2] provides a detailed description of charge transfer in conductor–insulator systems for various ratios of the conductivities of components. However, the real systems that are most often encountered in practice also include a third component—an oxide film on the conductor surface. Possessing a high resistance, this component can drastically change the situation (graphite and the powders of noble metals including gold, platinum, etc. are exceptions). Thus, when speaking of the percolation theory, one should take into consideration not only the "black" and "white" regions [2], but also the "gray" ones (representing oxides on the metal surfaces), which means the transition to a three-component system. The contact resistance R_c has been theoretically studied [3, 4] as a function of various parameters of the conductor particles (the maximum height of protrusions, the Young modulus, etc.) and of the contact pressure. The possible mechanisms of charge transfer between particles of the conducting phase and the possible fracture of the oxide film in polymer composites because of shrinkage of the binding material were also considered [3], but only in application to polymer composites. Investigation of the systems of other types, for example, the molecular crystal–conductor system, may reveal new features, which would enable one to rise to a higher level of generality in the problem under consideration. Despite a large number of papers devoted to percolation theory [1–3], questions concerning the value of the percolation threshold in different systems and its dependence on the contact resistance between the conductor particles generally remain open.

In this context, we have studied the dependence of the percolation threshold x_c on the contact resistance, the dependence of the bulk resistivity ρ of a macrosystem on the volume fraction x of the conducting phase, and the possibility of estimating the contact resistance in experiment.

EXPERIMENTAL

As an example of a disordered macrosystem, we will consider the paraffin–conductor system. Recently [5], it was demonstrated that the most technologically convenient insulators are low-melting paraffin and ceresin, which have comparatively high resistivity, can be easily shaped, and are chemically passive with respect to most metals [6]. As conductors, particles of graphite, iron, and aluminum were used. Guided by the above reasoning and favoring simplicity of the fabrication of samples, we chose thermographite (representing a colloid-graphite dry substance S-1 with a basic particle size of 4×10^{-6} m), a solid oil paraffin P1, pure-grade iron powder with a basic particle size of 6×10^{-6} m, and pure-grade aluminum powder with a basic particle size of 10×10^{-6} m after sifting.

The resistance of the oxide shells surrounding the iron and aluminum grains was estimated from the following considerations. It is known that the oxide film on an iron surface has the chemical composition $Fe_2O_3 \cdot nH_2O$ [6, 7] and can vary in resistance. In this study, the value of ρ for the iron oxide was experimentally determined for pure-grade powdered Fe_2O_3, which was stored under the same conditions as the iron. These measurements yielded the value $\rho = 1 \times 10^5 \, \Omega$ m, which was assumed to be equal to the resistivity of the shell on an iron particle. In air, aluminum forms a thin but fairly dense, stable, and high-ohmic film, which finds wide application in technology [8]. The value $\rho = 1 \times 10^{9.5} \, \Omega$ m experimentally determined for reagent-grade powdered aluminum oxide was used as the resistivity of the oxide shell on aluminum grains.

The resistivity of paraffin (matrix material) [5] was $\rho = 1 \times 10^{10.5} \, \Omega$ m. For minimizing the contact resistance, the use of graphite as the conducting phase [5] is advantageous in comparison to metals. Carbon oxides are gases and, whatever the origin of graphite, one may disregard the formation of oxide film on these particles.
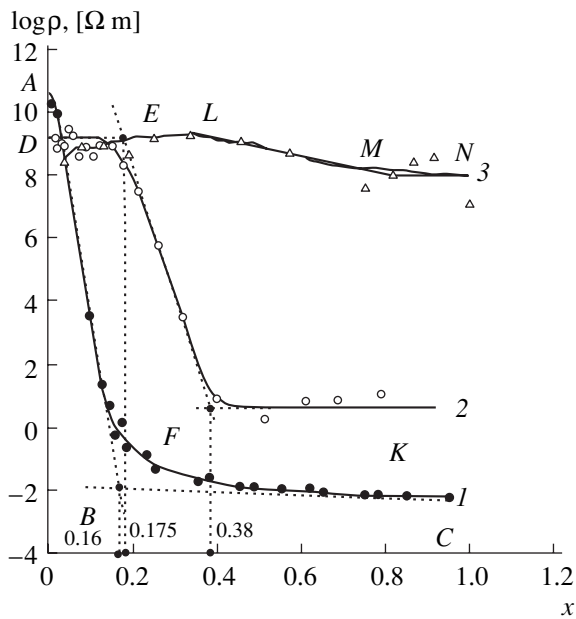
Plots of the bulk resistivity vs. volume concentration of the conducting phase for (*1*) paraffin–graphite, (*2*) paraffin–iron, and (*3*) paraffin–aluminum macrosystems.

Furthermore, carbon is chemically passive with respect to most insulators in a wide temperature range [6].

The composite samples had the shape of parallelepipeds with dimensions $(10 \times 10 \times 15) \times 10^{-3}$ m. The technology of melt preparation was as follows. Parameters such as the melt temperature, the rate and the time of stirring, etc., were chosen so that the final specimens with the same concentration of conducting phase would exhibit maximal (3–5%) spread in the electric parameters: resistance, capacitance, and quality factor. Special attention was paid to ensure the absence of air bubbles, separation from electrodes, and other defects. Prior to the preparation of melt, the iron powder was demagnetized. The electrodes were made of electrolytic copper. The electrodes were poured with liquid mixture of the components and occurred inside the specimen, except for a small contact area. No less than five specimens with the same concentration of the conducting phase were prepared for each concentration of the conducting phase. The bulk resistivity was measured using the conventional two-electrode scheme.

RESULTS AND DISCUSSION

Thus, we have three macrosystems involving conductors covered by oxide films with resistivities increasing in the following order: paraffin–carbon with $\rho \approx 0$, paraffin–iron with $\rho \approx 1 \times 10^5\ \Omega$ m, and paraffin–aluminum with $\rho \approx 1 \times 10^{9.5}\ \Omega$ m. Obviously, the contact resistance between particles depends on many factors [3], including the matrix shrinkage or, more strictly speaking, the shrinkage stress [4].

Let us estimate the shrinkage for paraffin. Assuming $\beta \approx 3\alpha$ and $\alpha = 130 \times 10^{-6}\ K^{-1}$ [9] (where β and α are the volumetric and linear expansion coefficients of paraffin, respectively), using the expression

$$\beta = \frac{1}{\Delta T}\frac{\Delta V}{V_0},$$

and taking $\Delta T \approx T_{melt} - T_{exp} \approx 65$ K, we estimate the shrinkage as

$$\frac{\Delta V}{V_0} = \Delta T \beta \approx 2.5\%.$$

This value of shrinkage is lower than the typical quantities for contactols [4]. Therefore, taking into account the high plasticity of paraffin [8, 9], one may assume that the shrinkage stresses in macrosystems with paraffin are small and the contact resistances form a series that is similar to the series of resistivities of the conductor grain shells.

The figure shows the curves of $\log\rho = f(x)$ for the three systems studied: paraffin–carbon, paraffin–iron, and paraffin–aluminum. All curves decrease monotonically. The most pronounced bending is observed for the paraffin–carbon system. The percolation threshold x_c was determined as the point of intersection of straight lines AB and BC, which approximate the steep and low-resistance portions of the curve, respectively. For the paraffin–carbon system, this method yields $x_c = 0.16$, which correlates with the data [4] and agrees well with the theory [1, 2]. The character of the $\log\rho = f(x)$ curve for the paraffin–iron composition is different, specifically, the plot actually consists of three portions: DE is a high-resistance portion (with x ranging from 0 to 0.175), where the resistance is almost independent of the conducting phase concentration and is determined by the resistivity of the matrix; EF is the portion of maximal change in ρ ($0.175 \le x < 0.38$), where the composite resistivity is already not strongly dependent on the resistivity of the matrix but is not yet determined by that of the conductor; and FK is the portion of minimal resistance, where the dependence $\log\rho = f(x)$ is weakly pronounced since ρ is governed by the resistivity of the conductor ($0.38 \le x < 1$). For aluminum (see figure), the percolation threshold is even more smeared in the range $0.45 \le x < 0.775$. Concerning the appearance of the plots, it seems reasonable that the percolation threshold of real systems should be understood as an interval rather than a point. Such an interval presents a range along x axis where the resistance typical of a high-ohmic matrix gives place to that typical of a conductor.

Let us now approximate the dependences in the ranges of the most dramatic change in the resistivity for all three compositions under consideration. For the paraffin–iron system, the most abrupt change in the conductivity corresponds to the EF branch ($0.175 \le x <$

0.38) and can be approximated by

$$\log\rho = 15.4 - 37.1x, \quad (1)$$

$$\rho = 10^{15.4} \times 10^{-37.1x}, \quad (2)$$

where $\gamma = -37.1$ is the slope of the $\log\rho = f(x)$ curve ($\Delta\log\rho = 7.98$ and $\Delta x = 0.215$).

For comparison, the most abrupt portion of the curve for graphite is described by the function $\rho = 10^{12} \times 10^{-87.5x}$, so that $\gamma = -87.5$ [4]. For the paraffin–aluminum system, the region LM can be described as

$$\log\rho = 10.5 - 3x, \quad (3)$$

$$\rho = 10^{10.5} \times 10^{-3x}, \quad (4)$$

and $\gamma = -3$ ($\Delta\log\rho = 1$ and $\Delta x = 0.325$).

In what follows, we consider the special features of the $\log\rho = f(x)$ curves and introduce the value that can be used as an indirect characteristic of the contact resistance. As was noted in [3], the data on R_c in real systems are missing from the literature on the subject, which is explained by the serious difficulties encountered in the experimental determination of R_c. In the classical form [1, 2], the percolation theory can be applied only to the paraffin–graphite system, since this theory implies that the mechanical touch of particles provides a reliable electrical contact. As was noted above, the situation in real systems is more complicated. Let us consider the percolation model for the paraffin–iron system, that is, in the case when ρ of the matrix is higher than that of the shell. When two conductor particles (iron–iron) in this (or any other similar) system come in touch, their contact resistance may change considerably, since the oxide shell of an iron grain is nonuniform over the surface area and, as a consequence, the contact resistance of particles upon touching can be different. In a disordered system, a random spread in the barrier height must strongly affect the situation. The point is that the conductivity of such a system as a whole is defined by barriers with a height that is close to the percolation threshold [3], that is, by the highest barriers. At $x_c \geq x$, the percolation threshold corresponds to the resistance of the most extended regions of the matrix (paraffin) in the forming conducting chain; while, at $x \geq x_c$, when the metal chain has already formed, it is the maximal contact resistance in the iron–iron chains through the oxide. Such barriers are far apart from each other and are connected by low-resistance regions; therefore, they experience an appreciable drop of voltage across them. As a result, the nonohmic character of conduction should be more pronounced in a random system than in an ordered one [3], where the potential barriers between the conductor particles are equal. The deviation of the bending point from the value 0.15 [1] can be reasonably explained by assuming that an excess concentration of iron is spent for the formation of branches parallel to the percolation barriers. In view of the stochastic character of the connection between particles, these branches may have a lower barrier height and a lower resistance. Thus, the "bypass channels" formed in the range $0.175 \leq x < 0.38$ eliminate high percolation barriers in the conducting chains. The higher the percolation thresholds and the greater their number, the less pronounced the percolation transition is, that is, the smaller the difference $\Delta\rho$ between the final values of resistivity and the wider the range Δx within which this change occurs. The ratio of the resistivity difference $\Delta\rho$ to the concentration change Δx just corresponds to the slope $\gamma = \Delta\log\rho/\Delta x$ of the $\log\rho = f(x)$ curve. Since our conduction model (supported by experimental results) predicts a clear decrease in the value of $\Delta\log\rho$ and an increase in Δx with increasing R_c, one may infer that $R_c = f(\gamma)$ and the value of γ can be used to characterize R_c.

It should be noted that, according to [3], there are various percolation thresholds in the polymer film–aluminum composites obtained using different technological processes, such as powder technology, polymerization filling, mixing of components in the polymer melt (rolling). Taking into consideration significant differences in the coefficients of linear thermal expansion of graphite (7×10^{-6} K^{-1}), iron (11.3×10^{-6} K^{-1}), aluminum (22.58×10^{-6} K^{-1}), and paraffin (130×10^{-6} K^{-1}) [9], we may suggest that shrinkage in the polymer–conductor system may result in much higher contact pressures [3, 8], which may destroy even a rather strong oxide shell of aluminum grains. High contact pressures may be related to the mechanical characteristics of polymers upon solidification, which are significantly different from those of paraffin. This is in line with the assumptions made at the beginning of this study.

CONCLUSIONS

(1) The dependence of the resistivity of the paraffin–conductor system on the conducting phase concentration has been experimentally studied, and it is shown that, for the oxide shells of particular conductors (graphite, iron, and aluminum with $\rho \approx 0$, 1×10^5, and $1 \times 10^{9.5}$ Ω m, respectively), the $\log\rho = f(x)$ curves vary in shape from an abrupt step for the paraffin–graphite composite to smeared transition for the paraffin–aluminum composite.

(2) Based on the experimental data obtained, a model of conduction in insulator–conductor systems with various contact resistances of the conducting phase was formulated in terms of the percolation theory. It is suggested that the contact resistance R_c can be indirectly characterized by the ratio of the final difference of the resistivity $\Delta\log\rho$ to the concentration range Δx, that is, by the value $\gamma = \Delta\log\rho/\Delta x$, which can be determined as the slope of the $\log\rho = f(x)$ curve in the percolation interval.

REFERENCES

1. J. M. Ziman, *Models of Disorder: the Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979).

2. B. I. Shklovskii and A. L. Éfros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).

3. I. A. Chmutin, S. V. Letyagin, V. G. Shevchenko, and A. T. Ponomarenko, Vysokomol. Soedin. **36**, 699 (1994).

4. V. E. Gul' and L. Z. Shenfen', *Conducting Polymer Compositions* (Khimiya, Moscow, 1984), p. 37 [in Russian].

5. V. A. Sotskov and S. V. Karpenko, Zh. Tekh. Fiz. **73** (1), 107 (2003) [Tech. Phys. **48**, 100 (2003)].

6. A. M. Petrov, Kh. V. Bal'yan, and A. T. Troshchenko, *Organic Chemistry* (Vysshaya Shkola, Moscow, 1963), p. 63 [in Russian].

7. *Chemical Encyclopedic Dictionary* (Sov. Éntsiklopediya, Moscow, 1983), p. 200 [in Russian].

8. V. V. Pasynkov, *Materials for Electronic Engineering* (Vysshaya Shkola, Moscow, 1980), pp. 155–160 [in Russian].

9. *Tables of Physical Data: Reference Book*, Ed. by I. K. Kikoin (Atomizdat, Moscow, 1976), p. 1005 [in Russian].

Translated by A. Sidorova

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Estimation of SEM Imaging Resolution of the Domain Structure of Pyroprobe-Heated Ferroelectrics

A. A. Sogr and A. G. Maslovskaya

Amur State University, Blagoveshchensk, 675027 Russia
e-mail: master@amursu.ru
Received November 27, 2003; in final form March 23, 2004

Abstract—The specific contrast of scanning electron microscope (SEM) images of ferroelectric domains observed in the pyrocurrent (pyroprobe) mode is analyzed. Calculations taking into account both the non-uniform heating of domains and the heat diffusion via the domain walls and the crystal boundaries are performed. It is established that the heat diffusion smears the domain images of small domains. Along with the probe diameter, the scan rate is shown to be an important factor determining the character of the SEM image contrast. A decrease in the scan rate may appreciably lower the resolution even in case of a fairly thin probe. © 2004 MAIK "Nauka/Interperiodica".

INTRODUCTION

The electron beam of a scanning electron microscope (SEM) is not merely a passive probe (indicator) of the geometric or potential profile of the sample surface under investigation, but produces ionizing, electric, and thermal action on the sample. The SEM imaging of ferroelectric domains in a special electron-beam induced polarization current (EBPC) mode was described in [1, 2]. In this mode, a video signal is formed by the polarization currents of electrons injected into the surface layers of a sample at high accelerating voltages. In [1], the local character of the electric action produced by the injected charge and the related resolution of this method were analyzed. The observed "shadow" effect is a typical piece of evidence of the domain response to the thermal action of the probe (pyroresponse). However, the diffusion of heat should considerably expand the heat-affected region and, thus, smear the domain images of small domains.

Previously, a focused electron beam probe was used as a local heat source for SEM imaging due to the pyroelectric effect [3, 4]. The pyroeffect that resulted from the action of a scanning thermal probe on a ferroelectric crystal was calculated by Latham [4], who considered the propagation of heat from a resting point probe in a plane-parallel plate of finite thickness and determined the pyroelectric potential induced in the plate in the open-circuit regime. The pyroelectric signal from a probe scanning over a domain structure was calculated as a superposition of signals from various parts of the probe on domains with opposite signs. This approach implied the local action of each part of the probe on the corresponding domain without allowance for the heat diffusion into adjacent domains.

The SEM image of a domain structure observed in the EBPC mode [1] is governed by a number of factors, both of electrical and thermal nature. The aim of this study is to separate and analyze the factors limiting the resolution of SEM imaging of ferroelectric domain structure in the pyroprobe mode.

STATEMENT OF THE PROBLEM AND THE METHOD OF SOLUTION

A local change in the polarization P of a plane short-circuit homogeneous plate with the thickness d_0 gives rise to the polarization current in the external circuit [5]:

$$I = \frac{1}{d_0}\int_V \frac{\partial P}{\partial t}dV = \frac{\gamma}{d_0}\int_V \frac{\partial T}{\partial t}dV, \tag{1}$$

where $\gamma = \partial P/\partial T$ is the pyrocoefficient, T is the temperature, and V is the sample volume.

The pulsed pyroresponse of a point domain with a pyrocoefficient in the form of the delta function is determined by the local rate of change in its temperature (that is, by the longitudinal component of the temperature gradient) and by the velocity of the probe:

$$I(t) = \gamma\frac{\partial T}{\partial t} = \gamma\frac{\partial T}{\partial x}v. \tag{2}$$

In our calculations, we relate the system of coordinates to a point source moving along the x axis and describe the stationary temperature field of this source in an infinite medium by the following equation [6]:

$$T(x', R) = \frac{W}{4\pi k\sqrt{x'^2 + R^2}}\exp\left(-\frac{v(x' + \sqrt{x'^2 + R^2})}{2a^2}\right), \tag{3}$$

where R and x' are cylindrical coordinates in the system related to the probe, v is the probe velocity, k is the ther-
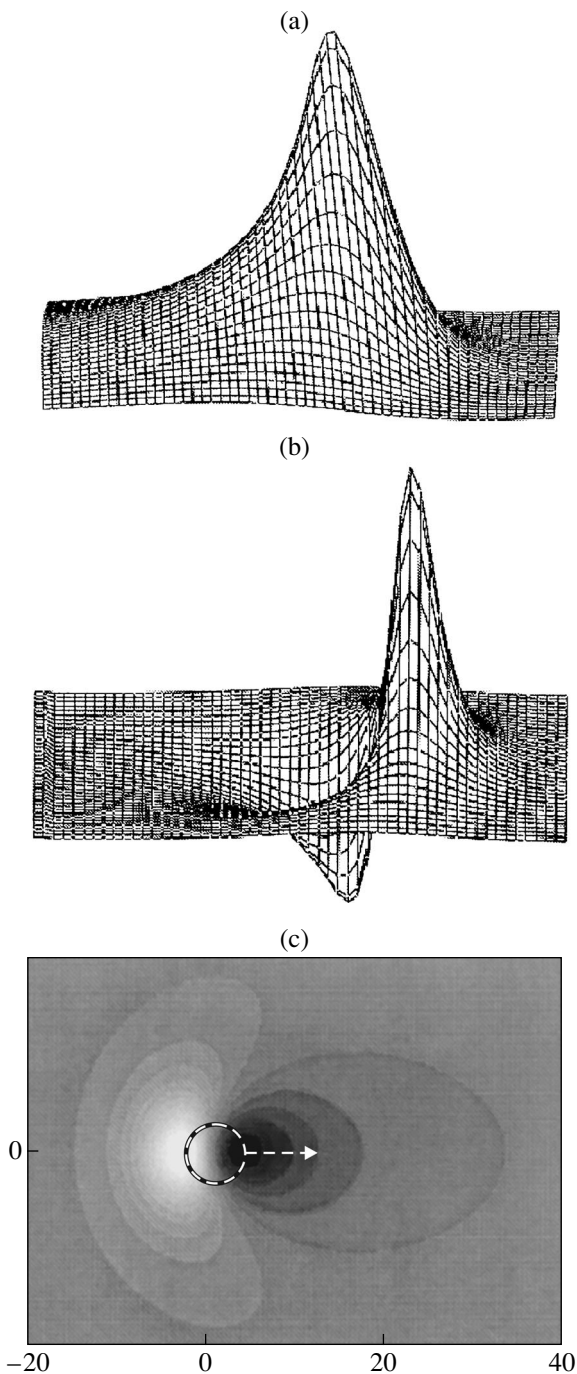
(a)

(b)

(c)

Fig. 1. The formation of a video signal from a point domain: (a) temperature profile of a moving point source; (b) profile of the longitudinal temperature gradient; and (c) calculated image of a point domain for the probe radius $R_p = 4$ (norm. units) and velocity $v = 0.25$ (norm. units).

mal conductivity, a^2 is the thermal diffusion factor, and W is the power of the probe.

This field differs from the stationary field of a point charge by the exponential factor, which produces compression of the leading edge of the isotherms ($x' > 0$).

The exponent parameter $r_0 = a^2/v$ is defined by the dimensionless Péclet parameter in terms of the heat conduction theory.

The boundary conditions for the thermal field allow the problem in a half-space to be considered in the infinite space with a doubled source power.

A pyrosignal formed by a point pyrodomain with x_0 coordinate reflects the derivative of the thermal field at the point x' (in the probe-related coordinate system) at a time moment given by the conventional expression $t = (x_0 - x')/v$. This corresponds to the point on the screen with the coordinate

$$x_s = v_s t = m(x_0 - x'), \qquad (4)$$

where v_s is the velocity of the beam motion over the screen and $m = v_s/v$ is the SEM image magnification factor.

As follows from (4), the longitudinal gradient of the thermal field is reflected on the screen with the inversion of coordinates.

Figure 1 shows the surface profile of (a) temperature and (b) its derivative with respect to coordinate x, which defines the SEM video signal from a point pyrodomain in a neutral matrix. Figure 1c presents the calculated image of a point domain produced by the motion of a finite-size probe. The dashed circle indicates the position of probe at the moment of crossing the domain (the origin of coordinates is placed at the point where the domain resides). The domain image appears as two peaks with opposite contrast.

In the case of a pyroactive matrix with a pyricoefficient of the opposite sign, the result will differ in the double amplitude and a constant equal to the pyroresponse of the matrix.

Due to the exponential temperature drop in front of the moving probe, the isotherms become additionally compressed at $r > r_0$ and, thus, the first peak is localized. The temperature gradient is maximum near the boundaries of the heat-evolving region; hence, the distance between the two peaks depends on the diameter of the heat probe ($2R_p$).

Behind the scanning probe, the temperature drops along the path according to the law $\sim x^{-1}$, the temperature gradient being proportional to $\sim x^{-2}$. Therefore, the signal in the tail decreases by one order of magnitude over a distance of about $3R_p$. The compression of the leading front is effective only at sufficiently high velocities, which correspond to $r_0 \approx R_p$; therefore, an increase in the scan rate will compress the first peak but the probe radius will still remain the decisive factor limiting the image size.

In SEM imaging, the size of the heat source, R_p, actually corresponds to the depth of electron penetration into the sample rather than to the diameter of the electron probe [7]. For triglycine sulfate (TGS) crystals

and an electronbeam energy of 15 keV, this value was estimated at 3 μm [1].

For an extended domain, superposition of signals from its different parts takes place. In particular, for a linear domain extended along the x axis ($x > 0$), the integration leads to the reconstruction of the temperature profile (Fig. 1a), so that the signal repeats the profile of the temperature rather than that of the temperature gradient.

In order to obtain the signal from a cylindrical domain with the radius r_d, the temperature gradient should also be integrated in the transverse direction. Figure 2 shows the profiles of a pyrosignal for the central line of cylindrical domains with various radii. As the probe approaches the domain edge, the step is smeared and the trailing edge of the signal drops, the latter effect decreasing with increasing domain radius.

Within the above-described approach to the calculation of pyroresponse, the point domain acts as a probe that "scans" over the thermal field in accordance with formula (4) in the direction opposite to that of the sweep. The calculation for a finite-size domain implies the superposition of signals from all points of the domain. This gives a qualitative insight into the basic principles of the image formation in the pyrocurrent mode.

Another approach to the calculation of video signal is based on the use of a nonstationary heat conduction equation [6]. Substituting an expression for $\partial T/\partial t$ from the nonstationary heat conduction equation [6] into Eq. (1) and using the Ostrogradsky theorem enables one to split the pyrosignal into two components [8]:

$$I = \frac{\gamma}{\rho c d_0} \int_V w \, dV + \frac{a^2 \gamma}{d_0} \oint_S (\mathrm{grad} T)_n dS, \qquad (5)$$

where ρ is the density of medium, c is the specific heat, w is the volume density of the source power, and S is the boundary of the region V.

Here, the first integral describes the pyrocurrent induced by the heat source in an unbounded medium. For a point source ($w = \delta(\mathbf{r}_p)$), where $\mathbf{r}_p$ is the radius-vector of the probe position) this term will exactly repeat the pyrocoefficient profile $\gamma(r)$. The diffusion of heat through the boundaries of a finite-size domain, which is described by the second integral, leads to a noticeable distortion of the initial (ideal) videosignal.

In our calculations, the values of x, t, and v were normalized with the help of the Fourier and Péclet criteria in accordance with the heat conduction theory; the physical parameters of the crystal and probe were normalized to the unit pyrocurrent value. In terms of the normalized variables, the compression factor r_0 relates to the velocity of motion as $r_0 = 1/v$.

The calculation of the pyrocurrent using formula (5) for the thermal field of a moving source (2) and a plane infinite boundary perpendicular to the direction of the
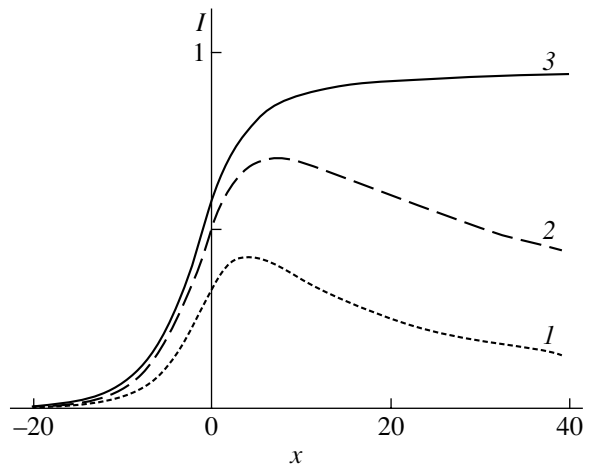

Fig. 2. Video signal profile from a cylindrical domain calculated for the probe radius $R_p = 4$ (norm. units); the domain radius $r_d = 10$ (*1*), 20 (*2*), and 80 (*3*); and scan rate $v = 0.25$ (norm. unites).
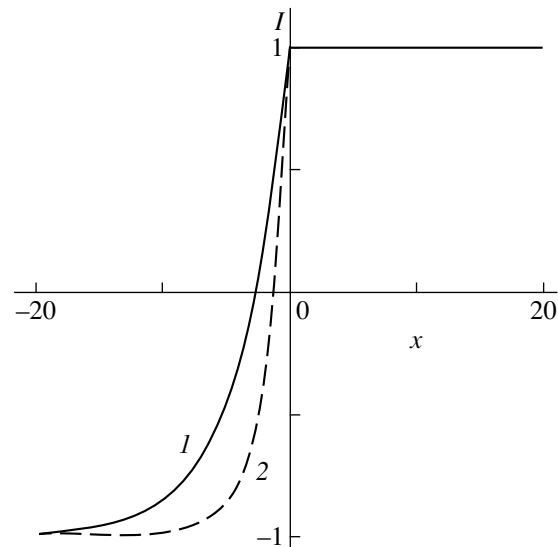

Fig. 3. Pyroresponse from a plane infinite boundary perpendicular to the direction of heat probe motion calculated for the velocities (*1*) $v_1 = 0.25$ and (*2*) $v_2 = 0.5$ (norm. units).

probe motion ($x = 0$) yields an exponential function characterized by the Péclet parameter r_0 and decreasing toward the beginning of sweep $x < 0$ (Fig. 3):

$$I(x) = (1 - \mathrm{sgn}(x)) \exp\left[\frac{x(1 - \mathrm{sgn}(x))}{2r_0}\right] + \mathrm{sgn}(x). \qquad (6)$$

The concept of splitting the pyrosignal into two components allows a drop of the signal at the edge of a cylindrical domain (Fig. 2) to be interpreted as the leakage of heat through its side surfaces. The signal decay time is governed by the domain radius via the Fourier number $t = r_d^2$ (in normalized variables); and the corre-

sponding length is equal to

$$l = vt = r_d^2/r_0. \qquad (7)$$

The drop of the signal grows at the domain edges. This also leads to a considerable distortion of the contrast of narrow extended domains.

RESULTS AND DISCUSSION

As follows from Eq. (6), the width of the transient region in the perpendicular wall image equals the compression factor r_0; therefore, for imaging a step without distortion, it is necessary to increase the scanning rate (Fig. 3). However, at $r_0 < R_p$, the image width is limited by the probe size, rather than by the compression factor, and further increase in the velocity is inexpedient. Note that, for a TGS crystal ($a^2 = 3 \times 10^{-7}$ m^2 s^{-1}) and the scan rate typical of SEM probe (scan time, 10 ms; scan path-length, 1 mm), the compression parameter is $r_0 = 3 \times 10^{-6}$ m, which is close to the size of the heat source. According to formula (7), an increase in the probe velocity also reduces the drop of signal in the narrow extended domains.

The application of this model to periodic domain structures with alternating transverse stripes of width h indicates that the contrast remains noticeable for the domain sizes $h \geq 1.5R_p$. With a decrease in the scan rate (and the corresponding increase in r_0), we also obtain the reduction and the following vanishing of the contrast. However, the contrast remains distinguishable even at a scan rate corresponding to $r_0 \approx 10h$, though the compression of the heat front within the domain width in this case appears to be quite insignificant.

CONCLUSIONS

The diffusion character of heat propagation imposes a serious limitation on the resolution of SEM imaging in the pyroprobe mode. The images of small domains are noticeably distorted. The contrast of extended domains with small transverse size decreases along the direction of the probe motion. Along with the probe size, the scan rate has also a pronounced effect on smearing the image of the transverse walls, though the latter effect is not as profound in a periodic domain structure.

REFERENCES

1. A. A. Cogr, Izv. Akad. Nauk, Ser. Fiz. **60**, 174 (1996).

2. A. A. Sogr and I. B. Kopylova, Ferroelectrics **191**, 193 (1997).

3. W. Clay, B. J. Evans, and R. V. Latham, J. Phys. D **7**, 1456 (1974).

4. R. V. Latham, J. Phys. D **9**, 2295 (1976).

5. V. M. Rudyak, *Switching Processes in Nonlinear Crystals* (Nauka, Moscow, 1986) [in Russian].

6. H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*, 2nd ed. (Clarendon, Oxford, 1959).

7. M. N. Filippov, Izv. Akad. Nauk, Ser. Fiz. **57**, 163 (1993).

8. A. A. Sogr and A. G. Maslovskaya, Izv. Akad. Nauk, Ser. Fiz. **67**, 1197 (2003).

Translated by A. Sidorova

BRIEF
COMMUNICATIONS

On the Role of Surface Energy in Nanodimensional Crystalline Objects

S. V. Karpenko and A. I. Temrokov

Research Institute of Applied Mathematics and Automation, Kabardino-Balkar Research Center,
Russian Academy of Sciences, Nalchik, 36000 Russia
e-mail: sv_karpenko@mail333.com
Received September 10, 2003

Abstract—The surface contribution to the thermodynamic potential for bulk and nanodimensional particles of ion crystals is estimated with the help of the electron statistical theory of ion crystal lattice. A number of size effects associated with the excess surface energy of ultradisperse particles are considered. In particular, the possibility of stability loss in the crystal lattice upon the transition of the surface energy to the range of negative values under a high pressure is predicted. © *2004 MAIK "Nauka/Interperiodica".*

Considerable advances in nanotechnologies [1, 2] have stimulated the development of appropriate analytic models for describing thermophysical properties of nanodimensional objects. The number of practical applications of the physics of nanodimensional systems is increasing and includes the application of ultradisperse catalysts of chemical reactions [3], description of nanocluster structures in the theory of contact melting [4], and obtaining of semiconducting heterostructures with quantum wires and quantum dots [5, 6]. The latter trend is associated with the study of a new class of materials constituted by macroscopic ensembles of small particles with a size ranging from 1 to 100 nm. The basic physical properties of such systems differ substantially from the properties of materials in the conventional bulk state and are unique in some cases. Most characteristics of nanoobjects are determined by the properties of individual small particles in the ensemble. In spite of considerable advances in the field of nanotechnologies, most available theoretical models are based on the assumption that the basic dynamic, thermodynamic, and mechanical characteristics of nanoobjects coincide with their values obtained in macroscopic experiments. However, for structures consisting of a few atomic layers, the contradiction between the obvious discreteness of the object and the continual nature of the method for its description in unavoidable.

When the particle size becomes commensurate with the characteristic correlation scale of a physical phenomenon, various size effects are observed in the systems of such particles. Nanodimensional objects also exhibit most clearly all features of surface states, since the fraction of surface atoms in such system may be on the order of 10%. In addition, a developed surface affects the lattice and electron subsystems of particles, substantially modifying the spectra of various elementary excitations, which are sensitive to a change in the symmetry and in the boundary conditions. Obviously, we can expect that the share of the surface energy in the total thermodynamic potential of a nanosystem sharply increases.

This communication is aimed at estimation of the surface contribution to the thermodynamic potential of a system for bulk and ultradisperse particles in ionic crystals with a structure of the sodium chloride (NaCl) type and at analysis of specific size effects that are associated, in our opinion, with relatively excessive surface energy of nanoparticles as compared to bulk objects. All calculations are made in the framework of the electron-statistical theory of the ion crystal lattice [7–9].

It is well known that the surface energy in equilibrium is a strictly positive quantity [10]. With increasing external pressure, the surface energy decreases [10, 11]; it is quite possible that, at a certain pressure $p = p_{cr}$, the surface energy vanishes and acquires negative values upon a further increase in pressure. Such a state of substance is naturally unstable and may lead to breakdown of the sample with the formation of particles of various degrees of dispersion.

The formalism of the electron density functional method [7, 8, 12] makes it possible to calculate the values of pressure at which the surface energy vanishes. The calculations were made for ion crystals with a lattice structure of the NaCl type.

We write the thermodynamic potential of a crystal under pressure in the form

$$G = \sum_{k=1}^{7} N_k U_k(a_k R)$$

$$- V \frac{\partial}{\partial V} \left[\sum_{k=1}^{7} N_k U_k(a_k R) \right] - \frac{\alpha_\mu}{R} + 4\pi r^2 k\sigma, \quad (1)$$

where α_μ is the Madelung constant, R is the nearest-neighbor distance, $V = 2R^3$ is the unit cell volume in the $B1$ phase, $U(R)$ is the pair interaction potential, $a_k = R_k/R_0$ is the ratio of the radius of the kth coordination sphere to the radius of the first coordination sphere, N_k is the coordination number, σ is the surface energy, and k is a numerical coefficient taking into account the deviation of the crystal shape from a sphere.

In accordance with the Gibbs distribution, the excess surface energy density $\sigma(hkl)$ at absolute zero is given by

$$\sigma(hkl) = \sum_i \sum_{j=0}^{\infty} (W_j^{(i)} - W_\infty^{(i)}) n_j(hkl), \qquad (2)$$

where $\sigma(hkl)$ is the surface energy density of the (hkl) face, $W_j^{(i)}$ is the energy of a particle in the jth layer associated with the ith type of ion-ion interaction forces, $W_\infty^{(i)}$ is the same for the bulk of the crystal, and $n_j(hkl)$ is the number of particles in the jth plane per unit area.

Let us consider in greater detail the approximation used here. One of the merits of the Gibbs approach is the rigorous choice of the dividing plane, which is very important since the value of the surface energy density $\sigma(hkl)$ depends on the position of the interface in the crystal, as the density of particles is different at different planes. This fact was noted in [13]. Our further analysis of the surface energy will be based on the method developed in [14]. The essence of this method is that the crystal is divided into 2D meshes and summation in the expression for the surface energy density is carried out over the aggregate of such meshes.

In the zeroth approximation used here, approximate expression (2) assumes the form

$$\sigma(hkl) = n_0(hkl) \sum_i (W_0^{(i)} - W_\infty^{(i)}). \qquad (3)$$
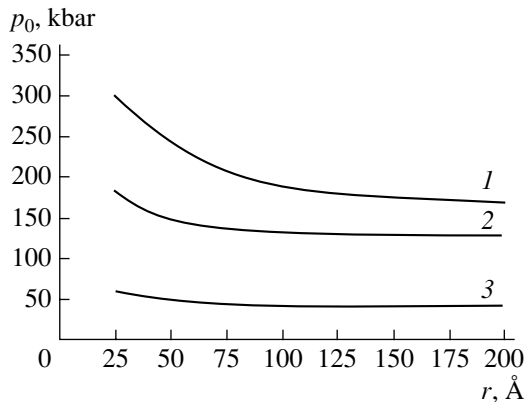
Dependence of the pressure corresponding to the $B1$–$B2$ phase transition in sodium halides on the crystal size: NaF (*1*), NaCl (*2*), and NaBr (*3*).

Let us consider a 2D mesh in an infinitely large solid. Obviously, for an undistorted crystal, we have

$$W_\infty^{(i)} = W_S^{(i)} + 2W_{v/2}^{(i)}, \qquad (4)$$

where $W_S^{(i)}$ is the energy of a particle on the mesh, which is associated with the ith type of the forces of interaction of this particle with all remaining particles in the given plane and $W_{v/2}^{(i)}$ is the energy of the same particle, which is determined by the interaction with all particles in all planes lying above or below the given plane.

Thus, the energy of a particle at the surface plane of an undistorted crystal is given by

$$W_0^{(i)} = W_S^{(i)} + W_{v/2}^{(i)}. \qquad (5)$$

Eliminating $W_{v/2}^{(i)}$ from Eqs. (4) and (5), we obtain

$$\sigma(hkl) = \frac{1}{2} n_0(hkl) \sum_i (W_S^{(i)} - W_\infty^{(i)}). \qquad (6)$$

We introduce the notation

$$\beta^{(i)} = \frac{W_S^{(i)}}{W_\infty^{(i)}} = \frac{A_S^{(i)}}{A_v^{(i)}}$$

for the ratio of the sums over the infinitely large 2D mesh and over an infinitely large lattice for the ith type of the ion-ion interaction forces. In this case, expression (6) assumes the form

$$\sigma(hkl) = \frac{1}{2} n_0(hkl) \sum_i (\beta^{(i)} - 1) W_\infty^{(i)}. \qquad (7)$$

In particular, for Coulomb forces, β is the ratio of the Madelung constants for a 2D mesh and for a 3D lattice. For the remaining forces, quantity β is the ratio of rapidly converging series that can be easily summed.

The calculation of the specific contribution of surface energy to thermodynamic potential (1) of the crystal leads to the following results: the surface energy for a bulk sample amounts to about 6% of the total energy; this value is equal to 24% for a small spherical crystalline particle of radius 100 Å and 68% for a particle of radius 50 Å (these data correspond to a sodium chloride crystal). Thus, the relative surface contribution to the thermodynamic potential sharply increases with decreasing particle size.

Having constructed thermodynamic potential (1), we can derive the required equation of state $p = -(dG/dV)$, which makes it possible to determine the distance between the particles for which the surface energy vanishes. Calculations performed for alkali-halide crystals give values of 1.17–1.21 for r_0/r_{cr}. Using the experimental results obtained in [15], we can easily obtain the value of pressure at which the surface energy vanishes from the pressure dependence of the lattice

Values of critical pressure for some alkali-halide compounds

Crystal	r_0/r_{cr}	p_{cr}, kbar
LiF	1.21	230
NaF	1.19	190
NaCl	1.18	130
NaBr	1.17	40
KCl	1.16	32
KBr	1.16	32

constant. The results of calculations for some alkali-halide crystals are given in table.

Analysis of the data compiled in the table shows that the energy of alkali-halide crystals becomes negative at pressures ranging from 30 to 230 kbar. The size effect of polymorphous $B1$–$B2$ transformation and metallization of insulators predicted by us for small-size crystals [7, 8] can also be classified as a size effect of the above type. A detailed description of this effect is given in [7]; the essence of the effect lies in an increase in the pressure of phase transformation upon a decrease in the crystal size. The figure shows the dependences of the pressure corresponding to polymorphic transformation on the particle size for sodium halides, which demonstrate a substantial increase in the phase transition pressure for nanodimensional crystals.

Summing up, it should be noted that the role of the surface energy for nanodimensional systems is much more significant than for bulk objects whose properties are mainly determined by the bulk contributions to the thermodynamic potential. The surface contribution for nanoparticles is comparable to the bulk contribution to the energy of the system, which determines the unique properties of such objects (in particular, the size effects described here). In our opinion, the most interesting is the possibility for existence of a state of the system with a negative surface energy. Such a state of matter is deliberately unstable. However, it remains unclear whether or not a sample with a negative surface energy will experience breakdown and what the breakdown mechanism is. The answer to this question can be obtained only after experimental investigations.

REFERENCES

1. Zh. I. Alferov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 3 (1998) [Semiconductors **32**, 1 (1998)].
2. A. M. Krivtsov and N. F. Morozov, Fiz. Tverd. Tela (St. Petersburg) **44**, 2158 (2002) [Phys. Solid State **44**, 2260 (2002)].
3. I. D. Morokhov, V. I. Petinov, L. I. Trusov, and V. F. Petrunin, Usp. Fiz. Nauk **133**, 653 (1981) [Sov. Phys. Usp. **24**, 295 (1981)].
4. J. J. Kim, H. A. Marzouk, C. C. Eloi, and J. D. Robertson, J. Appl. Phys. **78**, 245 (1995).
5. D. A. Vinokurov, V. A. Kapitonov, O. V. Kovalenkov, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 858 (1999) [Semiconductors **33**, 788 (1999)].
6. N. N. Ledentsov, V. M. Ustinov, V. A. Shchukin, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 385 (1998) [Semiconductors **32**, 343 (1998)].
7. S. V. Karpenko, A. Kh. Kyarov, A. I. Temrokov, and D. L. Vinokurskiĭ, Kristallografiya **47**, 326 (2002) [Crystallogr. Rep. **47**, 291 (2002)].
8. D. L. Vinokurskiĭ, S. V. Karpenko, A. Kh. Kyarov, and A. I. Temrokov, Dokl. Akad. Nauk **381**, 756 (2001) [Dokl. Phys. **46**, 835 (2001)].
9. A. V. Bushman and V. E. Fortov, Usp. Fiz. Nauk **140**, 177 (1983) [Sov. Phys. Usp. **26**, 465 (1983)].
10. A. I. Temrokov, Teplofiz. Vys. Temp. **38**, 573 (2000).
11. J. W. Gibbs, *The Scientific Papers*, Vol. 1: *Thermodynamics* (Longmans, Green, New York, 1906; Gostekhteorizdat, Moscow, 1950).
12. V. F. Ukhov, R. M. Kobeleva, G. V. Dedkov, and A. I. Temrokov, *Electron-Statistical Theory of Metals and Ionic Crystals* (Nauka, Moscow, 1982) [in Russian].
13. G. G. Benson and K. S. Yun, J. Chem. Phys. **42**, 3085 (1965).
14. S. N. Zadumkin and A. I. Temrokov, Izv. Vyssh. Uchebn. Zaved. Fiz. **9**, 40 (1968).
15. V. N. Zharkov and V. A. Kalinin, *Equations of State for Solids at High Pressures and Temperatures* (Nauka, Moscow, 1968; Consultants Bureau, New York, 1971).

Translated by N. Wadhwa

BRIEF COMMUNICATIONS

Exact Value of Conjugate Ion Current in an Electrolyte in the Gurevich–Kharkats Diffusion-Migration Model

A. E. Dubinov and I. D. Dubinova

*Federal State Unitary Enterprise Russian Federal Nuclear Center, All-Russia Research Institute
of Experimental Physics, Sarov, Nizhegorodskaya oblast, 607200 Russia*
e-mail: dubinov@ntc.vniief.ru
Received March 29, 2004

Abstract—The exact solution of the problem of ion transport current in an electrolyte in the presence of a conjugate flow of solvent is obtained in the framework of the one-dimensional Gurevich-Kharkats diffusion-migration model; the conditions for the existence of the solution are analyzed. © *2004 MAIK "Nauka/Interperiodica".*

The problem of ion transport in an electrolyte in the presence of a conjugate solvent flow was analyzed in [1]. The problem was solved using the following one-dimensional diffusion-migration model:

$$\frac{dC_1}{dx} - C_1\frac{d\psi}{dx} - \frac{v}{D_1}C_1 = \frac{j}{\text{Fa}D_1}, \qquad (1)$$

$$\frac{dC_2}{dx} + C_2\frac{d\psi}{dx} - \frac{v}{D_2}C_2 = 0, \qquad (2)$$

$$C_1 = C_2, \qquad (3)$$

where C_1 and C_2 are the anion and cation concentration, x is the coordinate, ψ is the electric potential, v is the constant velocity of the solvent flow, D_1 and D_2 are the diffusion coefficient for cations and anions, j is the current density, and Fa is the Faraday number.

The solution to the problem under the assumption that the solvent flow velocity is directly proportional to the current density ($v = \alpha j$) in normalized form for $C = C_1 = C_2$ is given by

$$\tilde{C}(\xi) = \left(1 + \frac{1}{2\beta}\right)\exp[\beta J(\xi - 1)] - \frac{1}{2\beta}, \qquad (4)$$

where the following notation has been introduced: $\tilde{C} = C/C_0$, $J = j/j_0$, $\beta = (\alpha j_0 L/2)(D_1^{-1} + D_2^{-1})$, $j_0 = \text{Fa}D_1C_0/L$, and $\xi = x/L$, L being the characteristic length of the problem (e.g., the thickness of the Nernst diffusion layer) and C_0, the value of concentration C at $\xi = 1$.

If we assume that the current of an electrochemical reaction is proportional to the ion concentration for $\xi = 0$ (i.e., $J = k\tilde{C}(0)$, where $k > 0$), we arrive at the following transcendental equation for J:

$$\frac{J}{k} = \frac{1}{2\beta}[(2\beta + 1)\exp(-\beta J) - 1]. \qquad (5)$$

This equation was derived in [1] and was subsequently analyzed only qualitatively. However, this equation has an exact solution, which can be written in the form

$$J = \frac{1}{\beta}\left\{W\left[\frac{k}{2}(2\beta + 1)\exp\frac{k}{2}\right] - \frac{k}{2}\right\}, \qquad (6)$$

where $W(x)$ is the so-called Lambert W function, which is the inverse function of $y = x\exp x$.

This function has been introduced in the apparatus of mathematical physics quite recently [2]. Examples of solutions of various problems in mathematical physics with the help of this function are given in [3–5].

The validity of the solutions obtained here can be verified by substituting it into Eq. (5).

By way of an example, we consider the curves $J(\beta)$ for various values of k (see figure); considerable computational expenditures would be required to obtain these curves without a knowledge of solution (6).

Let us briefly analyze solution (6). It can easily be seen that the value $\beta = -1/2$ is the vertical asymptote of dependence $J(\beta)$ for any k. In the absence of a solvent flow ($\beta = 0$), the value of current, viz., the ordinate of the points of intersection of the curves with the vertical axis, has the form

$$J(0) = 2\left\{1 - \left[1 + W\left(\frac{k}{2}\exp\frac{k}{2}\right)\right]^{-1}\right\}, \qquad (7)$$

it is equal to unity for $k = 2$ and two for $k \longrightarrow \infty$. The
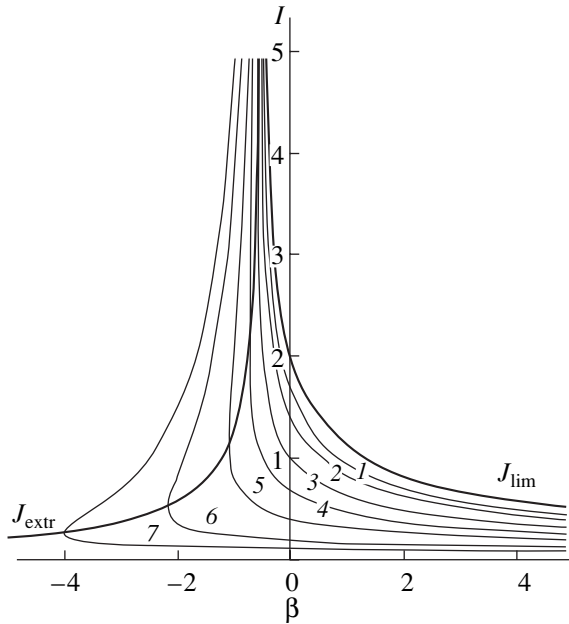
Dependence $J(\beta)$ for various values of k: 10 (*1*), 5 (*2*), 2 (*3*), 1 (*4*), 1/2 (*5*), 1/5 (*6*), and 1/10 (*7*); bold curves correspond to $J_{\text{lim}}(\beta)$ and $J_{\text{extr}}(\beta)$.

limiting current can be evaluated as the limit

$$J_{\text{lim}} = \lim_{k \to \infty} \frac{1}{\beta}\left\{ W\left[\frac{k}{2}(2\beta + 1)\exp\left(\frac{k}{2}\right)\right] - \frac{k}{2}\right\}$$

$$= \frac{1}{2}\ln(2\beta + 1). \tag{8}$$

It was mentioned in [1] that dependence $J(\beta)$ is double-valued for small values of k. Analysis of its stability shows that only the lower branch of the curve, which corresponds to the smaller value of the current, can be realized in actual practice. Consequently, it would be interesting to derive the equation for the curve, which is the locus of points of conjugation of the lower and upper branches of the plots for different values of k (equation of the extremal curve). In fact, the extremal curve and limiting current curve (8) define the bound-

aries of possible regimes of the model [1]. We can derive the equation for the extremal curve if we recollect that the Lambert function $W(x)$ also has two branches, the coordinates of the conjugation point being $(-1/e, -1)$. Then we obtain the following equation for the extremal curve in parametric form in k:

$$\begin{cases} \beta_{\text{extr}} = -\dfrac{1}{2}\left[1 + \dfrac{2}{k}\exp\left(-\dfrac{k}{2} - 1\right)\right], \\[3mm] J_{\text{extr}} = \dfrac{k(k+2)}{k + 2\exp\left(-\dfrac{k}{2} - 1\right)}. \end{cases} \tag{9}$$

Expressing k from the first relation in (9) and substituting it into the second relation, we arrive at the explicit dependence $J_{\text{extr}}(\beta_{\text{extr}})$:

$$J_{\text{extr}} = -\frac{1}{\beta_{\text{extr}}}\left[1 + W\left(-\frac{\exp(-1)}{2\beta_{\text{extr}} + 1}\right)\right]. \tag{10}$$

The curve for the limiting current and extremal curve are also shown in the figure.

Thus, we have obtained the exact solution to the problem of the ion transport current in an electrolyte in the presence of a conjugate solvent flow in the framework of the one-dimensional Gurevich-Kharkats diffusion-migration model [1] and have analyzed the boundaries of the existence of the solution.

REFERENCES

1. Yu. Ya. Gurevich and Yu. I. Kharkats, Dokl. Akad. Nauk SSSR **303**, 890 (1988).
2. R. M. Corless, G. H. Gonnet, D. E. J. Hare, *et al.*, Adv. Comput. Math. **5**, 329 (1996).
3. S. R. Valluri, D. J. Jeffrey, and R. M. Corless, Can. J. Phys. **78**, 823 (2000).
4. V. S. Bulygin, Elektrichestvo, No. 7, 68 (2002).
5. A. E. Dubinov, I. D. Dubinova, and S. K. Saykov, Dokl. Akad. Nauk **394**, 767 (2004) [Dokl. Phys. **49**, 132 (2004)].

Translated by N. Wadhwa

BRIEF
COMMUNICATIONS

Pulse Transmission through an Interface between Linear and Resonance Media

S. Sh. Tadjimuratov

Physicotechnical Institute, Academy of Sciences of Uzbekistan, Tashkent, 700084 Uzbekistan
e-mail: tadjimuratov@yahoo.com
Received December 1, 2003

Abstract—The problem of transmission of an optical pulse through an interface between linear and resonance media is addressed. Parameters of the transmitted and reflected waves versus the parameters of the incident wave are obtained. The amplitude and speed of the transmitted wave for an incident wave in the form of a soliton are calculated. © *2004 MAIK "Nauka/Interperiodica".*

1. As is well known, the parameters of waves reflected from and transmitted through an interface between two media can be calculated in the general form only when the media are linear. If one of the media is nonlinear or both of them are, it is necessary to consider each particular case individually (refer, e.g., to [1]).

The main difficulty in solving this problem in the general form is that functional forms of the waves in the two media are different and that the dependence of the polarization of a medium on the field is rather complex. As a result, the boundary conditions do not yield simple relationships for the wave parameters. However, if the equation that describes the evolution of a wave in a non-linear medium can be integrated by the inverse scattering transform method (ISTM), an opportunity arises to overcome these difficulties. In particular, the problem of pulse transmission through a resonance film that separates two linear media was solved in [2] based on the ISTM by introducing an additional fictive field with subsequent reduction of the equations to an integrable form. The problem in which both media are nonlinear was addressed in [3, 4], where the transition radiation of a soliton was calculated under the assumption that the media have similar parameters. In [5], it was proposed to analyze the wave boundary dynamics using the linear approach. Indeed, since distances in the boundary layer are small, it may be considered that the nonlinear effects have no time to develop and wave parameters can be calculated within the linear approximation.

In this study, we consider a pulse incident from a linear medium on the planar interface with a resonance medium consisting of two-level atoms. It is known [6] that such a medium can be described by a system of equations integrable by the ISTM. A relationship for the parameters of the electromagnetic waves is obtained under the assumption that their envelopes are smooth. Since the polarization of the medium depends on the field in a complex manner, a functional equation arises instead of a simple algebraic equation. To calculate the parameters of the transmitted and reflected waves at the interface, a self-consistent procedure is proposed, which involves relationships of the inverse transform method.

2. For simplicity, we will consider the normal incidence of a linearly polarized wave on the interface. Let the *yOz* plane of the coordinate system coincide with the interface, the *z* axis be parallel to the wave electric vector $\mathbf{E} = (0, E, 0)$, and the *x* axis be directed toward the resonance medium. Then, the magnetic field will be directed along the *z* axis: $\mathbf{H} = (0, 0, H)$.

Under these assumptions, the Maxwell's equations yield the following boundary conditions for the electric and magnetic field components in the first and second media:

$$E_1(x = 0) = E_2(x = 0),$$

$$H_1(x = 0) - H_2(x = 0) = \frac{4\pi}{c}\frac{\partial \tilde{P}_0}{\partial t}, \quad (1)$$

where $\tilde{P}_0$ is the surface polarization at the boundary of the resonance medium and c is the velocity of light in free space.

Let us represent the waves in the linear and resonance media and the polarization of the nonlinear medium as follows:

$$E_1 = \frac{1}{2}[E_0(x, t)\exp[i(k_1 x - \omega t)]$$

$$+ E_r(x, t)\exp[-i(k_1 x + \omega t)] + \text{c.c.}],$$

$$E_2 = \frac{1}{2}[E_t(x, t)\exp[i(k_2 x - \omega t)] + \text{c.c.}], \quad (2)$$

$$\tilde{P}_0 = \frac{1}{2}[P_0(x, t)\exp[i(k_2 x - \omega t)] + \text{c.c.}].$$

Here, E_0, E_r, E_t, and P_0 are the smooth envelopes of the incident, reflected, and transmitted waves and the polarization, respectively; ω is frequency; and k_1 and k_2 are the wave numbers in the linear and resonance media, respectively. Excluding the magnetic field from Eqs. (1), we obtain the following expressions for the smooth envelopes of the fields and the polarization:

$$E_0 + E_r = E_t,$$

$$k_1(E_0 - E_r) - k_2 E_t = -4i\pi\frac{\omega^2}{c^2}P_0. \tag{3}$$

If the resonance medium consists of two-level atoms, the polarization can be written as [6]

$$P_0 = -i n_0 p_0 \langle\lambda\rangle, \tag{4}$$

where n_0 is the surface concentration, p_0 is the dipole moment of atoms, the angle brackets mean averaging over the Doppler broadening, λ is a product of the amplitudes of the wave functions ϕ_1 and ϕ_2 of the ground and excited atomic states, respectively:

$$\lambda = -2\phi_1\phi_2^*, \tag{5}$$

where the asterisk means complex conjugate and ϕ_1 and ϕ_2 are to be found from the equations

$$\frac{\partial\phi_1}{\partial\tau} + i\eta\phi_1 = \frac{1}{2}\phi_2\varepsilon,$$

$$\frac{\partial\phi_2}{\partial\tau} - i\eta\phi_2 = -\frac{1}{2}\phi_1\varepsilon^*. \tag{6}$$

Here, $\eta = \Delta\omega/e\Omega$, $\Delta\omega$ is the frequency detuning due to the Doppler effect, $\Omega^2 = 2\pi n_0 p_0\omega/\hbar$, $\varepsilon = (p_0/\hbar\Omega)E_t(\xi, \tau)$, $\xi = \Omega e_2^{1/2}(x/c)$, $\tau = \Omega(t - e_2^{1/2}x/c)$, and $e_2^{1/2}$ is the linear part of the permittivity of the resonance medium. The wave in the resonance medium satisfies the equation

$$\frac{\partial\varepsilon}{\partial\xi} = \langle\lambda\rangle. \tag{7}$$

Equations (6) and (7) constitute a closed system, which describes the evolution of the electric field ε in the resonance medium. As was noted above, one can apply the ISTM [6] to these equations and solve the Cauchy problem, the initial condition for which is determined by Eqs. (3). Let us exclude the reflected wave from Eqs. (3) to express the transmitted wave in terms of the incident wave and polarization:

$$\varepsilon(0, \tau) = \gamma\varepsilon_0 + \sigma\langle\lambda\rangle. \tag{8}$$

Here, $\varepsilon_0 = (p_0/\hbar\Omega)E_0$, $\gamma = 2e_1^{1/2}/(e_1^{1/2} + e_2^{1/2})$ and $\sigma = 2p_0\Omega/(e_1^{1/2} + e_2^{1/2})c$. It is seen from Eq. (8) that, as was noted above, the transmitted wave depends not only on the incident wave but also on polarization, which, in turn, depends on E_t in a complex manner (see Eqs. (5), (6)). Further, we proceed as follows: calculate $\langle\lambda\rangle$ from

the zeroth-approximation functions ϕ_0 according to (5), substitute $\langle\lambda\rangle$ into (8) to refine the value of $\varepsilon(0, \tau)$, and use the refined value to find the function ϕ in the first approximation. Next, we iterate this self-consistent procedure to successively refine ε.

Accordingly, we substitute (8) into (6):

$$\frac{\partial\phi_1}{\partial\tau} + i\eta\phi_1 = \frac{1}{2}\phi_2(\varepsilon_0 + \delta\varepsilon),$$

$$\frac{\partial\phi_2}{\partial\tau} - i\eta\phi_2 = -\frac{1}{2}\phi_1(\varepsilon_0^* + \delta\varepsilon^*), \tag{9}$$

where $\varepsilon_0 = \gamma(p_0/\hbar\Omega)E_0$ and $\delta\varepsilon = -\sigma(p_0/\hbar\Omega)\langle\lambda\rangle$.

We will not solve (9) directly. Instead, regarding ϕ as a functional of E and calculating δE as a variation of the potential, we find a change in ϕ as a variation. Then, the complete solution will have the form

$$\phi = \phi_0 + \int\limits_{-\infty}^{\infty}\left[\frac{\delta\phi}{\delta\varepsilon}\delta\varepsilon + \frac{\delta\phi}{\delta\varepsilon^*}\delta\varepsilon^*\right]d\tau. \tag{10}$$

Here, ϕ is the column vector $\phi = (\phi_1, \phi_2)^T$, the superscript T means transposition, and ϕ_0 is the solution at $\delta\varepsilon = 0$. As ϕ, we take the function with the asymptotic behavior

$$\lim_{\tau\to-\infty}\phi(\tau)\exp(-i\eta\tau) = \begin{pmatrix}0\\-1\end{pmatrix}. \tag{11}$$

The variational derivatives can be found by varying Eqs. (6) (see, e.g., [7]):

$$\frac{\delta\phi}{\delta E} = \frac{\theta(\tau-\tau')}{2\tilde{a}}\phi_2(\tau')[\psi_2(\tau')\phi(\tau) - \phi_2(\tau')\psi(\tau)],$$

$$\frac{\delta\phi}{\delta E^*} = \frac{\theta(\tau-\tau')}{2\tilde{a}}\phi_1(\tau')[\psi_1(\tau')\phi(\tau) - \phi_1(\tau')\psi(\tau)], \tag{12}$$

where $\tilde{a}$ is the Jost coefficient, which is used in the inverse transform method, and $\theta(\tau)$ is the step function

$$\theta(\tau) = \begin{cases}1 & \text{for} \quad \tau > 0\\0 & \text{for} \quad \tau < 0.\end{cases}$$

The function ψ is another solution to (8), which is determined by the asymptotic behavior

$$\lim_{\tau\to\infty}\psi(\tau)\exp(i\eta\tau) = \begin{pmatrix}1\\0\end{pmatrix}. \tag{13}$$

Substituting (12) into (10) yields the following expression for ϕ:

$$\phi = \phi_0 + \frac{\sigma}{2\tilde{a}}[\phi_0 I_1 - \psi_0 I_2]. \tag{14}$$

Here,

$$I_1 = -2 \int_{-\infty}^{\tau} [\phi_2(\tau')\psi_2(\tau')\langle\phi_1\phi_2^*\rangle + \phi_1(\tau')\psi_1(\tau')\langle\phi_1^*\phi_2\rangle]d\tau',$$

(15)

$$I_2 = -2 \int_{-\infty}^{\tau} [\phi_2^2(\tau')\langle\phi_1\phi_2^*\rangle + \phi_1^2(\tau')\langle\phi_1^*\phi_2\rangle]d\tau'.$$

Now, we can calculate $\langle\lambda\rangle$ and, consequently, the parameters of the transmitted and reflected waves. Equations (3), (5), (8), and (14) solve this problem in the general form. As an example, consider the case when the incident pulse is a soliton:

$$\varepsilon_0 = \frac{4\beta}{\gamma}\mathrm{sech}(2\beta t).$$

(16)

The solution to system of equations (6) at $\delta\varepsilon = 0$ is

$$\phi_0 = \frac{1}{\eta - i\beta}\begin{pmatrix} i\beta\,\mathrm{sech}(2\beta t) \\ -[\eta + 1\beta\tanh(2\beta t)] \end{pmatrix}.$$

Calculating $\langle\lambda\rangle$ and substituting it into (8), we obtain the expression for the transmitted wave

$$\varepsilon(0,t) = \frac{4\beta}{\gamma}\mathrm{sech}(2\beta t)\left[1 + \frac{\sigma\beta\Delta_0}{2}\tanh(2\beta t)\right.$$

$$-\frac{\sigma^2\beta^2\Delta_0^2}{2}\mathrm{sech}^2(2\beta t) + \frac{\sigma^2\beta^2\Delta_0^2}{4} - \frac{\sigma^2\beta\delta_1^2}{4} - \frac{\sigma^3\beta^3\Delta_0^3}{8}$$ (17)

$$\left.+i\left[\frac{\sigma\Delta_1}{2} + \frac{\sigma^2\beta\Delta_0\Delta_1}{2}\tanh(2\beta t) - \frac{\sigma^2\beta\Delta_0^2\Delta_1}{8}\mathrm{sech}^2(2\beta t)\right]\right].$$

Here,

$$\Delta_0 = \left\langle\frac{1}{\eta^2 + \beta^2}\right\rangle, \quad \Delta_1 = \left\langle\frac{\eta}{\eta^2 + \beta^2}\right\rangle.$$

(18)

It can be seen that, after passing through the interface, the pulse acquires a phase shift and its shape changes. As the pulse described by (17) moves further, it may produce a single-soliton or a multisoliton pulse. Assuming that the condition for the formation of a single-soliton pulse is satisfied, let us calculate its parameters. To this end, within the ISTM, it is necessary to solve the Zakharov–Shabat eigenvalue problem, which has the same form as (6) with η replaced by the spectral parameter k:

$$\frac{\partial\varphi_1}{\partial\tau} + ik\varphi_1 = \frac{1}{2}\varphi_2\varepsilon,$$

$$\frac{\partial\varphi_2}{\partial\tau} - ik\varphi_2 = -\frac{1}{2}\varphi_1\varepsilon^*.$$

(19)

The complex value of k determines the speed and amplitude of the pulse produced. As we did previously when solving system (9), we calculate the change in k as a variation, which gives the change in the parameters

of the soliton as a result of its transmission through the interface. We obtain the following expressions for the differences in the amplitudes and speeds of the incident and transmitted solitons:

$$\Delta\frac{4\beta}{\gamma} = \frac{2\beta^3\sigma^2\Delta_0^2}{3\gamma} + 2\beta\sigma^2\Delta_1^{2\gamma},$$

$$\Delta\nu = \frac{2\beta^2\sigma^2\Delta_0}{3} + \frac{2\sigma^2\Delta_1^2}{\Delta_0} - \frac{16\beta^2\sigma^2\Delta_1^2}{3\Delta_0}.$$

(20)

It can be seen that, after passing through the interface, the soliton amplitude increases. Note that, if the Doppler broadening is an even function, the expressions for the amplitude and speed are simplified:

$$\Delta\frac{4\beta}{\gamma} = \frac{2\beta^3\sigma^2\Delta_0^2}{3\gamma},$$

$$\Delta\nu = \frac{2\beta^2\sigma^2\Delta_0}{3}.$$

(21)

Thus, in this study, expressions for the profile and parameters of the transmitted and reflected waves are derived for a wave passing from a linear to a resonance medium. The profile and parameters of the transmitted wave are calculated for an incident pulse in the form of a soliton. Only this case is considered for brevity; other cases should not present any difficulties.

ACKNOWLEDGMENTS

REFERENCES

1. B. B. Boǐko and N. S. Petrov, *Reflection of Light by Amplifying and Nonlinear Media* (Nauka i Tekhnika, Minsk, 1988) [in Russian].

2. V. I. Rupasov and V. I. Yudson, Zh. Éksp. Teor. Fiz. **93**, 494 (1987) [Sov. Phys. JETP **66**, 282 (1987)].

3. F. Kh. Abdullaev and R. Dzhangiryan, Zh. Tekh. Fiz. **53**, 2307 (1983) [Sov. Phys. Tech. Phys. **28**, 1418 (1983)].

4. Y. S. Kivchar and B. A. Malomed, Rev. Mod. Phys. **61**, 763 (1989).

5. F. Kh. Abdullaev, S. A. Darmanyan, and P. Bussimer, in *Proceedings of the Workshop "Optical Solitons," Singapore, 1990*, Ed. by F. Kh. Abdullaev (World Sci., Singapore, 1990), pp. 13–20.

6. G. L. Lamb, Jr., *Elements of Soliton Theory* (Wiley, New York, 1980; Mir, Moscow, 1983).

7. V. I. Karpman and V. E. Maslov, Zh. Éksp. Teor. Fiz. **73**, 537 (1978) [Sov. Phys. JETP **48**, 252 (1978)].

Translated by A. Khzmalyan

BRIEF
COMMUNICATIONS

On the Possibility of Transverse Size Oscillations in a Domain with a High Free-Electron Concentration under the Action of a Short Light Pulse on a Semiconductor

M. M. Loginova and V. A. Trofimov

Moscow State University, Vorob'evy gory, Moscow, 119992 Russia
e-mail: vatro@cs.msu.su
Received December 24, 2003

Abstract—Computer simulation is used to demonstrate the possibility of transverse size oscillations in a domain with a high concentration of free electrons under the action of a light pulse on a nonlinearly absorbing semiconductor. The results of computer simulation are confirmed analytically by testing the system stability. © 2004 MAIK "Nauka/Interperiodica".

The action of a laser pulse on a semiconductor induces optical bistability (OB) in the dependence of the absorption coefficient on the free charge carrier concentration. Under certain conditions, OB can be accompanied by the development of self-oscillations of free electron concentration, which are induced by an increase in the beam intensity [1], beam focusing to the bulk of the medium [2], diffraction effects [3], and the effect of self-induced electric field [4] on a high-absorption domain. Obviously, oscillations of the absorption coefficient of a semiconductor may lead to loss of information recorded in an OB element. Consequently, it is important to study this regime for practical applications. It should be emphasized that, in our opinion, the latter mechanism of excitation of oscillations due to the effect of light-induced electric field is most interesting, since it takes place for short pulses, which are predominantly used to elevate the speed of switching from a state of an OB element.

It should be noted that the oscillations of a high-absorption domain detected in [4] and obtained for a stronger dependence of the absorption coefficient on the concentration of free electrons generated from a donor level (see below) were characterized by a change of the region of transparency and strong absorption in the illuminated part of the beam and were observed upon a change in the electron mobility in a certain range of its values and in a small interval of the parameter characterizing the maximal concentration of free electrons. Here, we describe a new scenario of oscillations of the width of a domain with a high concentration of free electrons. This scenario is characterized, first, by manifestations of oscillations for zero electron mobility as well, which is equivalent to an increase in the duration of the acting pulse (the electric field becomes uniform). Second, oscillations of the width of a domain with a high concentration of free charge carriers occur upon a slow decrease in the free electron concentration

in the vicinity of the light beam center. Third, several regions of high gradients of free electron concentration (regions of a large space charge) exist in the unilluminated region of the medium.

It is well known [5] that, in the approximation of an optically thin layer, the process of interaction of a short laser pulse with a semiconductor can be described by the system of dimensionless differential equations

$$\frac{\partial^2 \varphi}{\partial x^2} = \gamma(n - N),$$

$$\frac{\partial n}{\partial t} = D\frac{\partial}{\partial x}\left(\frac{\partial n}{\partial x} - \mu n\frac{\partial \varphi}{\partial x}\right) + G(N, \varphi) - R(n, N),$$

$$0 < x < L_x = 1, \quad t > 0, \tag{1}$$

$$\frac{\partial N}{\partial t} = G(\varphi, N) - R(n, N)$$

with the initial boundary conditions

$$\left.\frac{\partial \varphi}{\partial x}\right|_{x=0, L_x} = 0, \quad \left.\frac{\partial n}{\partial x}\right|_{x=0, L_x} = 0,$$

$$n|_{t=0} = N|_{t=0} = n_0. \tag{2}$$

Functions G and R describing the generation of electrons from the donor level and radiative recombination of free charges in the semiconductor will be defined as follows:

$$G = q_0 q(x) q(t)\delta(N, n), \quad R = \frac{nN - n_0^2}{\tau_p}. \tag{3}$$

In system of equations (1)–(3), the following notation is introduced: x is the dimensionless transverse coordinate normalized to the radius of the incident optical beam; t is the time measured in units of relaxation
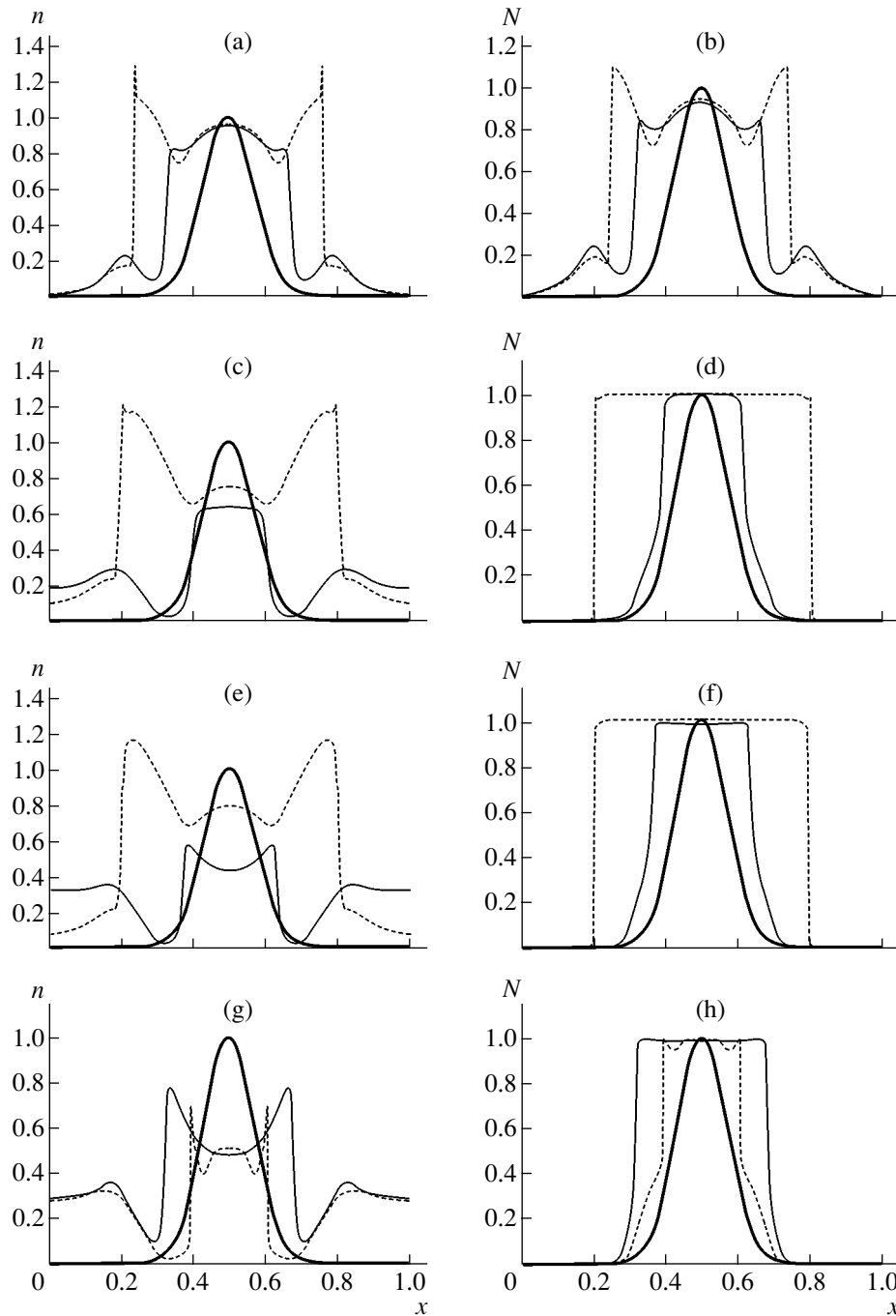
Fig. 1. Concentration distributions of free electrons and ionized donors observed for the interaction of a light beam with a semiconductor for values of parameters $D = 10^{-5}$, $\gamma = 10^3$, $\mu = 0$, $n_0 = 0.01$, $\psi = 2.553$, $\xi = 5$, and $q_0 = 1$ at time instants $t = 450$ (a), 550 (b), 1500 (c, d), 1250 (e, f) and 3750 (g, h) (dashed curves) and 500 (a), 600 (b), 2250 (c, d), 5000 (e, f), and 4000 (g, h) (solid curve). The bold curves correspond to the initial distribution of the input intensity of optical radiation.

time; and $n(x, t)$ and $N(x, t)$ are the concentrations of free electrons in the conduction band of the semiconductor and of ionized donors, normalized to their maximal possible values under the given conditions. Function $\varphi(x, t)$ is the dimensionless electric field potential, μ characterizes the electron mobility, and D is the diffusion coefficient for electrons. Parameter γ depends, in

particular, on the maximal possible concentration of free charge carriers; n_0 is the equilibrium value of concentration of electrons and ionized donors; and τ_p is the recombination time for free charge carriers. Function $q(x)q(t)$ describes the intensity profile and the temporal shape of an optical pulse with a maximal value of q_0. In

numerical experiments, a semiconductor was subjected to the action of a Gaussian beam with the maximal intensity

$$q(x)q(t) = e^{-\left(\frac{x - L_x/2}{0.1L_x}\right)^2}(1 - e^{-10t}) \qquad (4)$$

rapidly attaining its steady-state value.

The light energy absorption coefficient $\delta(N, n)$ can be approximated by the function

$$\delta(N, n) = (1 - N)e^{-\psi(1 - \xi n)}, \qquad (5)$$

which is close to one of the experimental dependences. Note that for certain values of parameters (e.g., $D = 10^{-5}$, $\gamma = 10^3$, $n_0 = 0.01$, $\mu = 0$, $\psi = 2.553$, and $\xi = 3$), an explosive increase in the concentration of free charges (and ionized donors) takes place upon an increase in the input intensity of optical radiation from $q_0 = 0.15$ to 0.2. This demonstrates the existence of OB in the optical radiation-semiconductor system. However, for a given intensity of the input pulse, inverse switching of free electron concentration to its lower level occurs after a certain time.

Apart from the formation of switching waves, an increase in parameter ξ leads to self-oscillations of the width of the band with a high concentration of free electrons and ionized donors (Fig. 1). At certain instants, the free electron concentration distribution profile is twice as large as the illuminated region width, while at other instants it coincides with the beam radius (Figs. 1c and 1d). The concentration of free electrons on the beam axis first attains its maximal value and then begins to decrease monotonically, while its value at the lateral faces of the crystal increases monotonically. In the vicinity of the beam center, the free electron concentration at certain instants either attains a local maximum on the beam axis, or a local minimum (Figs. 1e and 1g). The profile of concentration n at certain instants is close to the hyper-Gaussian distribution. However, at other instants, the distribution of the free electron concentration in the range $0.2 \leq x \leq 0.8$ acquires several local extrema located symmetrically relative to the beam axis (Figs. 1c, 1e, and 1g). The difference in the values of n at these points may be as large as 0.5.

Analogous oscillations of the width of the high-concentration region take place for the concentration of ionized donors. However, variations of the value of function N at the beam axis and at the crystal boundary are small and the difference between the values of function N at the points of local maxima and minima are much smaller than the corresponding difference for function n (Figs. 1d, 1f, and 1h).

We must separately consider the physical mechanism of realization of self-oscillations of the width of the high-concentration region. This mechanism involves a periodic variation of the absorption coefficient (in our case) in the vicinity of the boundary of the illuminated region. After high-concentration domains attain their minimal width (this corresponds to solid curves in Fig. 1), the absorption coefficient increases at the domain walls ($x \approx 0.3, 0.7$ at earlier instants and $x \approx 0.4, 0.6$ at later instants). This leads to generation of free charges in the vicinity of outer domain walls. Due to electron diffusion, the absorption coefficient also gradually increases in the range with an even lower intensity of optical radiation ($x \leq 0.3$ and $x \geq 0.7$). In this region, the absorption coefficient also increases and free charges are generated in spite of the low intensity of optical radiation. Further, on the one hand, the decrease in the number density of free charge carriers occurs at a higher rate as compared to the rate of generation of free electrons due to the increase in the recombination rate; as a consequence, the concentration of free charges decreases in this region. On the other hand, a part of free electrons is displaced to the crystal boundary due to diffusion. Since the concentration of ionized donors is low in this region, the recombination rate for free charge carriers is also low. This explains the slow increase in the free electron concentration near the crystal boundary.

The electron mobility μ substantially influences the form of the interaction due to the presence of narrow regions of space charge (see Figs. 1a, 2a). For example, an increase in the mobility first smoothens the distribution of free electron concentration and reduces the amplitude of self-oscillations of the transverse size of the region of high electron concentration. With a further increase in μ, the distribution of function n assumes the form of a monotonic curve; self-oscillations vanish, and the semiconductor-light beam system approaches its steady-state distribution quite rapidly. For example, for parameters $D = 10^{-5}$, $\gamma = 10^3$, $\mu = 10$, $n_0 = 0.1$, $\psi = 2.553$, $\xi = 3$, and $q_0 = 1$, this time is on the order of 100 dimensionless units. By way of an example, we consider Fig. 2, in which the free electron concentration distributions are shown for $\mu = 1$ and 10. It is clearly seen that the distribution of free electron concentration becomes smooth and close to the hyper-Gaussian distribution for large values of μ, while oscillations in the free electron concentration distribution are observed for small values of μ.

To verify the results of computer simulation, the stability of the solution to the initial problem was tested analytically in the vicinity of the steady-state distribution, where oscillations of the width of the high-concentration region are observed. A characteristic equation was constructed for a linearized system of relatively small perturbations of the functions in the vicinity of the virtually uniform axial distribution of concentrations. In accordance with the Hurwitz theorem [6], the necessary and sufficient conditions for stability of the concentration distribution for small perturbations are positive values of all principal diagonal minors of the Hurwitz matrix written for the coefficients of the characteristic equation of the system.
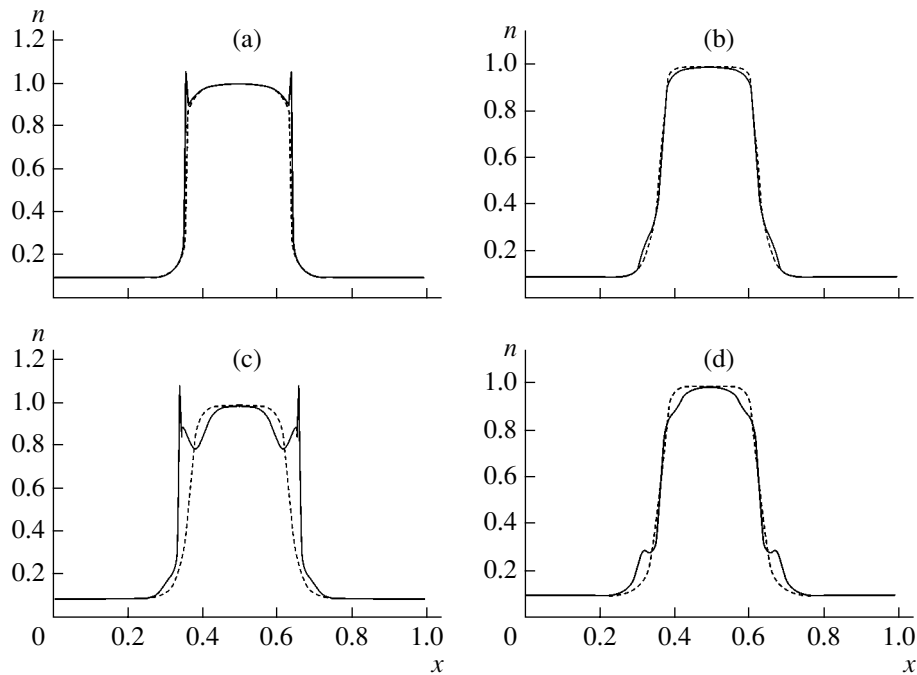
Fig. 2. Concentration distributions of free electrons observed for the interaction of a light beam with a semiconductor for values of parameters $D = 10^{-5}$, $\gamma = 10^3$, $n_0 = 0.1$, $\psi = 2.553$, $\xi = 3$, $q_0 = 1$, and $\mu = 1$ (solid curves) and $\mu = 10$ (dashed curves) at time instants $t = 10$ (a), 30 (b), 85 (c), and 100 (d).

These conditions were monitored for the above computer calculations. Analysis of the conditions proved that, for the concentration distributions shown in Figs. 1 and 2, stability was not observed near the beam axis. If, however, oscillations of the high-concentration region were absent in computer calculations, analytic expression also confirmed the stability of the given system. Thus, we can state that oscillating modes of variation of the concentration are inherent in the interaction between the light beam and the semiconductor studied here under the conditions of OB and light-induced electric field.

REFERENCES

1. H. Gibbs, *Optical Bistability: Controlling Light with Light* (Academic, New York, 1985; Mir, Moscow, 1988).
2. N. N. Rozanov, *Optical Bistability and Hysteresis in Distributed Nonlinear Systems* (Nauka, Moscow, 1997) [in Russian].
3. O. A. Gunaze and V. A. Trofimov, Pis'ma Zh. Tekh. Fiz. **23** (21), 6 (1997) [Tech. Phys. Lett. **23**, 846 (1997)].
4. O. S. Bondarenko and V. A. Trofimov, Pis'ma Zh. Tekh. Fiz. **22** (19), 6 (1996) [Tech. Phys. Lett. **22**, 779 (1996)].
5. R. A. Smith, *Semiconductors* (Cambridge University Press, Cambridge, 1959; Inostrannaya Literatura, Moscow, 1961).
6. L. Elsgolts, *Differential Equations and the Calculus of Variations* (Nauka, Moscow, 1969; Pergamon, Oxford, 1961).

Translated by N. Wadhwa

BRIEF COMMUNICATIONS

Experiments on the Injection of Dust Jets into Plasma

D. V. Vyalykh, A. E. Dubinov, I. L. L'vov, S. A. Sadovoĭ, and V. D. Selemir

All-Russia Research Institute of Experimental Physics, Russian Federal Nuclear Center,
Sarov, Nizhni Novgorod oblast, 607190 Russia
e-mail: dubinov@ntc.vniief.ru
Received February 18, 2004

Abstract—An experimental technique for studying the injection of dust jets into the plasma of a glow discharge in air is developed. The velocity and flight time of a dust jet are measured under different initial conditions. It is shown that the propagation of dust grains in plasma is accompanied by self-contraction instability along and across the discharge, which leads to dust agglomerization and clusterization. © *2004 MAIK "Nauka/Interperiodica".*

In dusty-plasma experiments, an important point is how the condensed-phase (dust) grains get into the plasma volume. First, dust grains can arise from the gas phase due to plasmochemical reactions [1] and condensation in a decaying plasma [2], or they can occur in the plasma because of the erosion of electrodes [3] and insulators [4]. Second, dust grains can be injected into plasma using special devices.

We analyzed various methods for injecting dust grains into plasma and the corresponding types of injectors, such as a mechanically shaken meshy dust container placed at the top of the chamber [5, 6]; the vibrating membrane of a loud speaker, which was set at the bottom of the chamber and threw the grains upward [7]; and a paddle-wheel dispenser rotating inside the chamber around a horizontal axis and lifting the grains up [8, 9].

We experimentally tested the first two methods. It was found that the above methods and devices did not provide the sufficient locality and accuracy of injection because of a significant scatter in the initial coordinates and velocities of the dust grains entering the plasma.

For this reason, we proposed a concept of a precise needle-type injector of dust grains, designed it, and experimentally verified its high performance. Using this injector, experiments with a 10-cm flight base were performed on the downward injection of dust grains into a discharge plasma. The grains, which were 60 ± 5 μm in size, fell freely across the discharge channel. It was found that the grains were efficiently scattered by collective plasma oscillations [10]. Using the needle injector, we also succeeded in orienting elongated dust grains with a diameter of 100 μm and length of 3 mm [11].

It was expected that the use of smaller dust grains would allow one to reveal new effects that did not manifest themselves in experiments with heavy grains. It turned out, however, that the vertical injection of grains through a needle under the action of gravitational force

was somewhat problematic: grains with a size of less than 40 μm significantly stuck to one another and choked up the needle channel. Therefore, we somewhat modified the experimental scheme (see Fig. 1) in order to inject small-size grains using the already developed needle injector. The experiment setup was modified as follows: A medical syringe in which the part of a needle with a bevel cut was cut off perpendicularly to the needle axis was used as a dust injector. The injector was set horizontally in such a manner that the needle went through a small hole in the cathode and the needle end was positioned flush with the cathode surface facing the discharge gap. The dust grains were injected along the horizontal axis of a dc glow discharge. Inside the syringe, there was a gas-discharge spark gap controlled by a BING-5 generator [12] of high-voltage pulses with
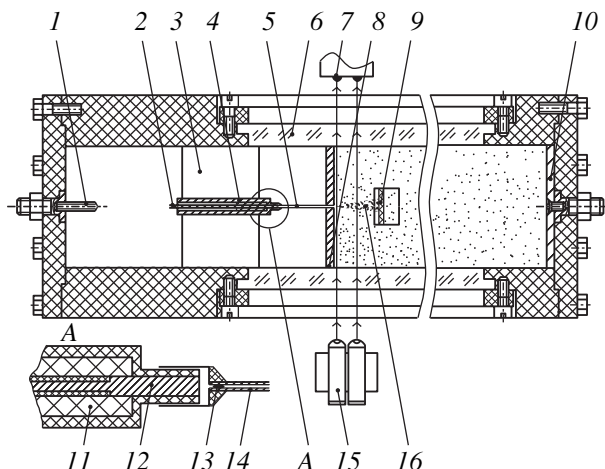

Fig. 1. Schematic of the experimental device: (*1*) connecting bus, (*2*) positive electrode connected to the output of the BING-5 generator, (*3*) rack, (*4*) dust injector, (*5*) needle, (*6*) view window, (*7*) photodetector, (*8*) cathode, (*9*) target, (*10*) anode, (*11*) insulator, (*12*) electrode, (*13*) dust grains, (*14*) needle channel, (*15*) laser, and (*16*) dust jet.
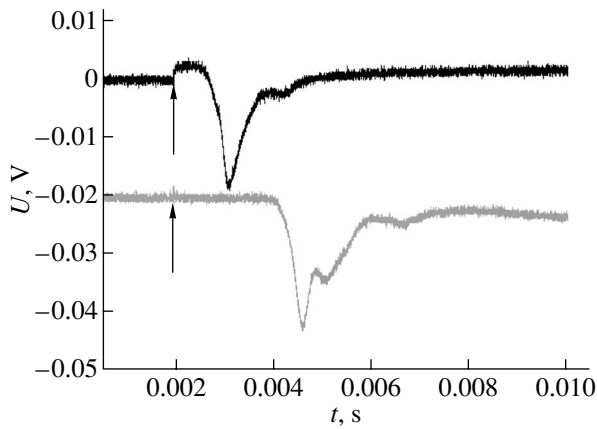
Fig. 2. Waveforms of the shadow pulses in air at a pressure of $P = 750$ Torr in the absence of a discharge (here and in the subsequent figures, the arrows show the instant of triggering a nanosecond spark discharge).
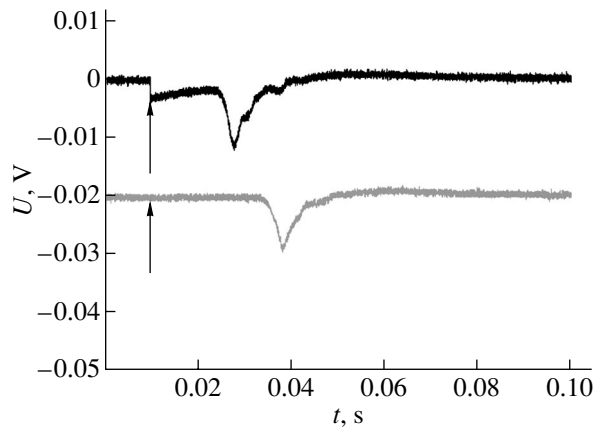

Fig. 3. Waveforms of the shadow pulses in air at a pressure of $P = 1$ Torr in the absence of a discharge.
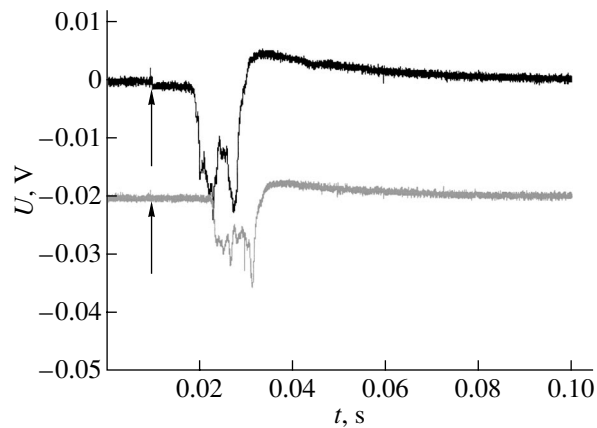

Fig. 4. Waveforms of the shadow pulses in air a pressure of $P = 1$ Torr in the presence of a glow discharge with a discharge current of $I = 0.1$ A (the electron density is $n_e = 2.6 \times 10^6$ cm^{-3}, and the electron temperature is $T_e = 5$ eV).

an amplitude of 10 kV and a rise time of 5 ns. After triggering a pulsed spark discharge inside the syringe, the arising shock wave kicked out the grains from the needle channel. A similar technique for injecting solitary grains into plasma was described in [13]. The above spark source of shock waves was tested in [14] at pressures of ~1 Torr and lower.

We developed a technology for filling the needle channel with approximately equal numbers (about 10^4) of dust grains. In brief, the needle was plunged to a definite depth into a grain container. Using a special ramrod with a diameter nearly the same as the needle channel diameter, a portion of the grains was then pushed from the needle end to the needle base.

The loaded injector was placed into a gas-discharge chamber with a 450-mm-long interelectrode gap and a 100×100 mm cross-sectional area of the chamber and electrodes. The chamber was then prepared for igniting a glow discharge. The parameters of the discharge plasma were determined using the probe technique. After the needed steady-state regime of a discharge was established, the dust grains were injected. In our experiments, we used dielectric (silicon carbide) grains 20 ± 5 μm in size and a 50-mm-long needle with a channel diameter of 200 μm. The length of the dust portion in the needle channel was 7 mm. The working gas was air at a pressure of 1 Torr.

The first series of experiments was aimed at measuring the velocity of dust jets. For this purpose, we used two cw red-light ($\lambda = 680$ nm) semiconductor lasers with an output power of 3 mW. The 3-mm-diameter laser beams traversed the expected path of the dust jet at distances of 5 and 15 mm from the needle end. On the opposite side from this path, there were two FD-256 photodetectors whose signals were fed to the inputs of a Tectronics TDS-3052B two-channel digital oscilloscope. In the absence of a dust jet, the oscilloscope displayed base lines. In the presence of a jet traversed by the laser beams, the photodetectors were shaded for a short time and the oscilloscope displayed pulsed signals (called below shadow pulses). A similar technique for detecting dust grains in plasma with the use of a single laser beam was implemented in [3].

Figure 2 presents typical waveforms of the shadow pulses in the case of injection into atmospheric-pressure air in the absence of a discharge. The estimated velocity of the dust jet front is 10 ± 1.5 m/s (hereinafter, the results are averaged over a series of ten pulses). In the case of injection into air at a pressure of 1 Torr, the estimated velocity of the dust jet front is 0.15 ± 0.02 m/s (Fig. 3). Such a great difference in the jet velocities at different pressures is quite understandable. Indeed, the intensity of the shock wave produced by the spark discharge decreases significantly with decreasing gas pressure. In both cases, however, the jet length was ~1 cm, being approximately the same over the entire flight base of the measurements. This length is a bit larger than the length of the initial portion of dust grains in the loaded needle channel.
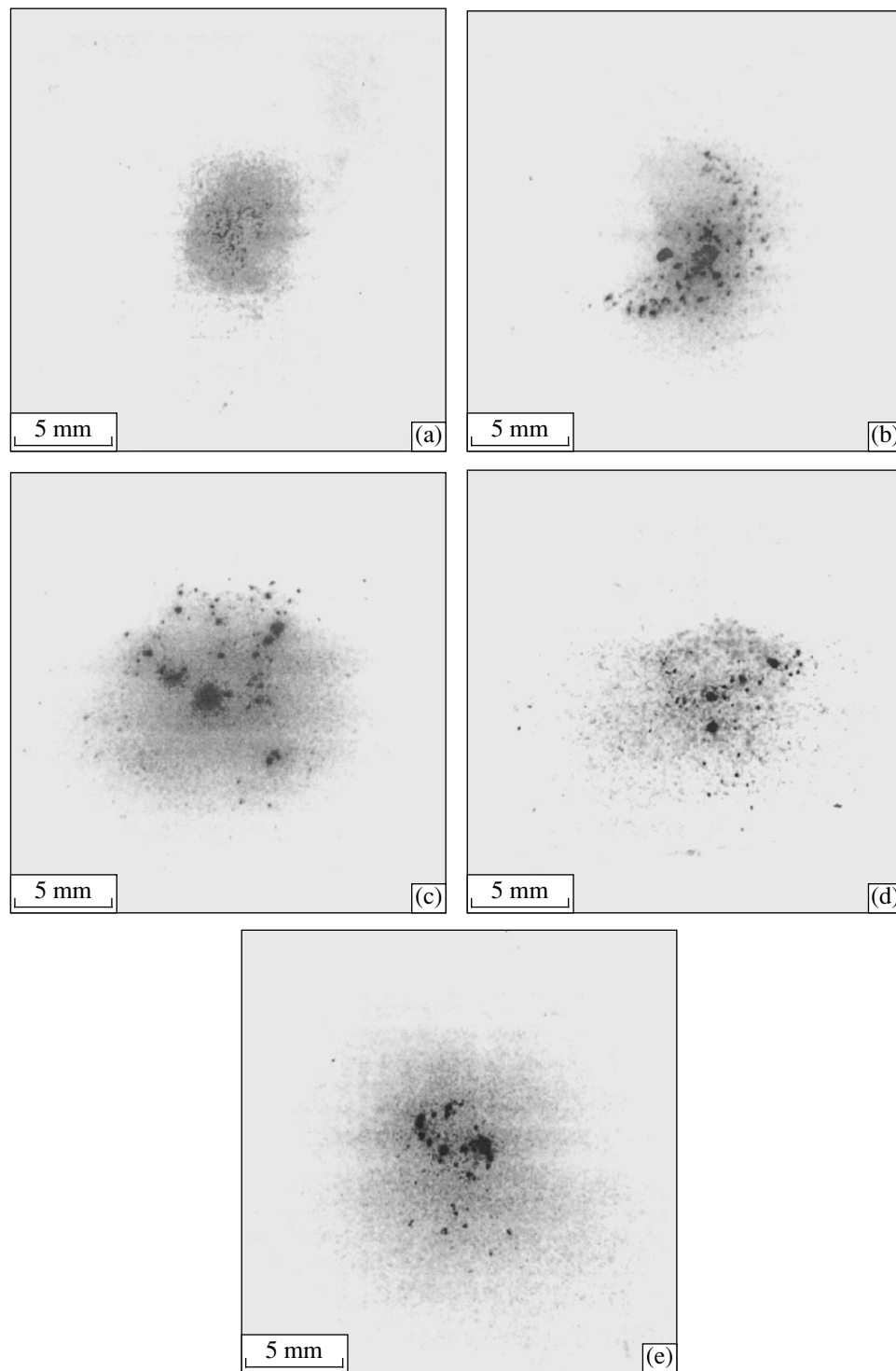
Fig. 5. Imprints of dust jets propagating in the plasma of a glow discharge in air at a pressure of $P = 1$ Torr and different discharge currents: $I =$ (a) 0, (b) 0.05, (c) 0.1, (d) 0.15, and (e) 0.25 μA.

Figure 4 presents typical waveforms of the shadow pulses in the case of injection into the plasma of a glow discharge with a current of 0.1 A (the corresponding electron density and electron temperature measured by the probe technique are $n_e = 2.6 \times 10^6$ cm^{-3} and $T_e = 5$ eV). It turned out that the velocity of the dust jet front in the plasma was 0.5 ± 0.08 m/s, which was significantly higher than that in a nonionized gas at the same pressure. A comparison of the waveforms in Figs. 4 and 5 show that the duration of the shadow pulses is approximately the same in both cases. This indicates that the dust jet in plasma undergoes acceleration as a whole

over a certain fraction of its path, rather than spreading out in the longitudinal direction. Estimates show that the charged dust grains can be accelerated by the electric field in the cathode sheath. Note that the waveforms in Fig. 4 are more irregular. This fact points to the onset of instability in the jet in the direction of its propagation, which leads to the dust agglomerization and clusterization. Presumably, this is the self-contraction instability described in [15, 16] and related to the intergrain attraction, which was earlier revealed by us in [17]. It follows from the average duration of the microspikes in Fig. 4 that the longitudinal size of the dust agglomerates is a few tenths of a millimeter or less.

The second series of experiments was devoted to studying the transverse structure of a dust jet. To this end, a 37×37-mm target was placed at a distance of 35 mm (i.e., at a distance longer than the jet longitudinal size) from the needle end. The target was a thin film with a glue layer facing the jet (we used conventional Rusi Star scotch tape). After taking an imprint, the film with captured grains was tightly (without folds) pasted on a Star Frost slide and then marked, after which it became possible to store it for a long time to be processed later. Figure 5a shows a typical imprint of a dust jet propagating in nonionized air at a pressure of 1 Torr. The image of the imprint was obtained by processing with an optical scanner with a 1200 dpi resolution. The shape of the imprint is nearly a circle with an azimuthally-uniform distribution of dust grains; the dust concentration slightly decreases from the center to the periphery.

The imprints of a dust jet propagating in a glow-discharge plasma at a pressure of 1 Torr and different discharge currents are shown in Figs. 5b–5e. It can be seen that the imprint size of the dust jet in plasma is somewhat larger than that in nonionized air (Fig. 5a); however, the imprint shape is still close to a circle. An increase in the size of the imprints in plasma as compared to that in a nonionized gas at the same pressure seems to be related to an additional scattering of dust grains by plasma oscillations.

An analysis of the imprints (Fig. 5) shows that the processes of the dust agglomerization and clusterization also take place across the jet. The characteristic transverse size of these dust agglomerates is from a few hundredths to a few tenths of a millimeter. Therefore, the agglomerates contain from a few tens to a few hundreds of dust grains.

The observed dust agglomerization and clusterization in plasma is a direct experimental evidence of attraction between likely charged grains. Such attraction was earlier observed only in one- and two-dimensional simulations [17]. Previously, the dust agglomerization and clusterization was observed only for nanometer grains (e.g., for silicon grains growing from the gas phase in an argon–silicon plasma [18]). To the best of our knowledge, the clusterization of grains with a size larger than 10 μm in plasma has been observed by us for the first time.

Thus, in this study, a technique for studying the injection of dust jets into the plasma of a dc glow discharge has been developed, the velocity of the dust jet front in plasma has been measured (it happens to be somewhat higher than that of a jet in a nonionized gas), and the longitudinal and transverse agglomerization and clusterization of dust grains in plasma (which is direct experimental evidence of attraction between likely charged grains) have been observed.

REFERENCES

1. M. Garrity, T. W. Peterson, and J. F. O'Hanlon, J. Vac. Sci. Technol. A **14**, 550 (1996).
2. B. M. Smirnov, Teplofiz. Vys. Temp. **31**, 176 (1993).
3. J. E. Jenkins and P. A. Chatterton, J. Phys. D **10**, L17 (1977).
4. V. L. Bychkov, A. Yu. Gridin, and A. I. Klimov, Teplofiz. Vys. Temp. **32**, 190 (1994).
5. V. E. Fortov, A. Nefedov, V. M. Torchinsky, *et al.*, Phys. Lett. A **229**, 317 (1997).
6. N. Y. Misconi, Laser Part. Beams **14**, 501 (1996).
7. T. Shoji, M. Fujigaya, and H. Tomita, Phys. Scr., T **89**, 41 (2001).
8. W. Xu, N. D'Angelo, and R. L. Merlino, J. Geophys. Res. **98**, 7843 (1993).
9. A. Barkan, N. D'Angelo, and R. L. Merlino, Phys. Rev. Lett. **73**, 3093 (1994).
10. S. V. Bulychev, A. E. Dubinov, V. S. Zhdanov, *et al.*, Prikl. Mekh. Tekh. Fiz. **42** (6), 19 (2001).
11. S. V. Bulychev, A. E. Dubinov, Yu. B. Kudasov, *et al.*, Pis'ma Zh. Tekh. Fiz. **29** (15), 45 (2003) [Tech. Phys. Lett. **29**, 636 (2003)].
12. B. G. Kudasov, S. S. Pavlov, V. A. Tananakin, *et al.*, in *Proceedings of the 11th IEEE International Pulsed Power Conference, Baltimore, 1998*, Ed. by G. Cooperstein and I. Vitkovitsky, p. 1572.
13. M. K. Asanaliev, Zh. Zh. Zheenbaev, M. A. Samsonov, and V. S. Éngel'sht, Fiz. Khim. Obrab. Mater., No. 3, 65 (1978).
14. A. E. Dubinov, S. A. Sadovoy, and V. D. Selemir, Shock Waves **10**, 73 (2000).
15. V. N. Tsytovich and D. Rezendes, Fiz. Plazmy **24**, 71 (1998) [Plasma Phys. Rep. **24**, 65 (1998)].
16. A. E. Dubinov, D. V. Selemir, and V. Sh. Shaĭdulin, Pis'ma Zh. Tekh. Fiz. **25** (17), 84 (1999) [Tech. Phys. Lett. **25**, 712 (1999)].
17. A. E. Dubinov, V. S. Zhdanov, A. M. Ignatov, *et al.*, Pis'ma Zh. Tekh. Fiz. **25** (13), 73 (1999) [Tech. Phys. Lett. **25**, 538 (1999)].
18. R. J. Buss and S. V. Babu, J. Vac. Sci. Technol. A **14**, 577 (1996).

Translated by N. Ustinovskiĭ

BRIEF
COMMUNICATIONS

Effect of Compressibility on the Stability of a Perfectly Conducting Plasma Jet of Plane and Cylindrical Geometry

I. A. Zhvania*, V. G. Kirtskhalia*, and A. A. Rukhadze**

* Sukhumi Institute of Physics and Technology, Sukhumi, 384914 Abkhazia, Georgia
e-mail: sipt@myoffice.ge
** Prokhorov Institute of General Physics, Russian Academy of Sciences,
ul. Vavilova 38, Moscow, 119991 Russia
Received March 10, 2004

Abstract—The effect of compressibility on the stability of plasma jets with boundaries (tangential discontinuity surfaces) of different configurations is studied. It is shown that, depending on the relationship between the MHD parameters of the jet and the surrounding medium, the compressibility may have a stabilizing or a destabilizing effect. It is also shown that, under certain conditions, the compressibility effect depends on the perturbation wavelength. © 2004 MAIK "Nauka/Interperiodica".

(1) In the MHD approximation, the natural oscillations of a perfectly conducting plasma jet are described by the dispersion relation [1–3]

$$\frac{N_2(\xi)}{\nu N_1(\xi)} = -G(\delta, \xi), \qquad (1)$$

where

$$N_1(\xi) = b_1^2 - \xi^2, \qquad (2)$$

$$N_2(\xi) = b_2^2 - (a - \xi)^2. \qquad (3)$$

The form of function $G(\delta, \xi)$ depends on the geometry of the jet boundary, which is a tangential discontinuity surface:

$$G(\delta, \xi) = \begin{cases} \dfrac{m_2}{m_1} \tanh(m_1 \delta) & \text{for a plane jet} \\[2ex] \dfrac{m_2 I_0(m_1 \delta) K_1(m_2 \delta)}{m_1 I_1(m_1 \delta) K_0(m_2 \delta)} & \text{for a cylindrical jet.} \end{cases} \qquad (4)$$

Here,

$$m_1^2 = \frac{(b_1^2 - \xi^2)(\mu_1^2 - \xi^2)}{\mu_1^2 b_1^2 - (\mu_1^2 + b_1^2)\xi^2} = 1 + \varepsilon_1(\xi), \qquad (5)$$

$$m_2^2 = \frac{[b_2^2 - (a - \xi)^2][\mu_2^2 - (a - \xi)^2]}{\mu_2^2 b_2^2 - (\mu_2^2 + b_2^2)(a - \xi)^2} = 1 + \varepsilon_2(\xi), \qquad (6)$$

$$\varepsilon_1(\xi) = \frac{\xi^4}{\mu_1^2 b_1^2 - (\mu_1^2 + b_1^2)\xi^2}, \qquad (7)$$

$$\varepsilon_2(\xi) = \frac{(a - \xi)^4}{\mu_2^2 b_2^2 - (\mu_2^2 + b_2^2)(a - \xi)^2}, \qquad (8)$$

and I_0, I_1, K_0, and K_1 are modified Bessel functions. The rest of the notation is as follows:

$$\xi = \frac{V - U_p}{C}, \quad a = \frac{V}{C}, \quad b_1 = \frac{V_{A1}}{C}, \quad b_2 = \frac{V_{A2}}{C},$$

$$\nu = \frac{\rho_1}{\rho_2}, \quad \mu_1 = \frac{C_1}{C}, \quad \mu_2 = \frac{C_2}{C}, \quad \delta = \frac{kd}{2} = \frac{\pi d}{\lambda},$$

where V is the jet velocity with respect to an immobile plasma medium, $V_{Ai} = H_i / \sqrt{4\pi\rho_i}$ is the Alfvén speed, ρ_i is the density, C_i is the speed of sound, C is a constant having the dimensionality of velocity, d is the thickness or the diameter of the jet, $U_p = \omega/k$ is the phase velocity, and λ is the wavelength of the perturbation specified in the form of a plane wave.

Dispersion relation (1) was derived under the assumption that the vectors $\mathbf{V}$, $\mathbf{k}$, and $\mathbf{H}_i$ point along the z axis, which is the symmetry axis of the cylindrical jet. The plane jet is assumed to be symmetric about the $x = 0$ plane, in which case we will be interested in the half-space $x \geq 0$. Hence, the perturbations are given by the following formulas:

$$f_i(x, z, t) = f_i(x) \exp[i(kz - \omega t)] \text{ for a plane jet, } (9)$$

$$f_i(r, z, t) = f_i(r) \exp[i(kz - \omega t)]$$
$$\text{for a cylindrical jet,} \qquad (10)$$

where the subscripts $i = 1$ and 2 refer to the regions inside and outside the jet, respectively. It is also
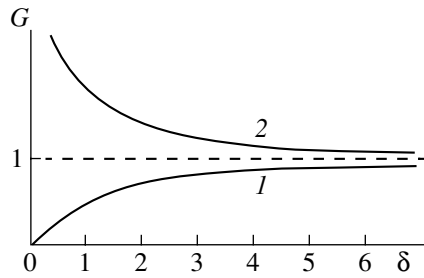
Fig. 1. Plots of the function $G(\delta)$ for (*1*) a plane jet and (*2*) a cylindrical jet.

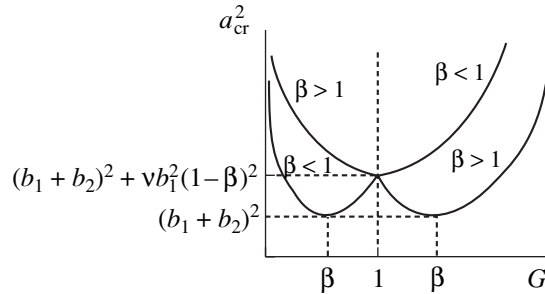**Fig. 2.** Dependence of a_{cr}^2 on $G(\delta)$ for an incompressible plasma jet: $G(\delta) < 1$ for a plane jet, and $G(\delta) > 1$ for a cylindrical jet.

assumed that the waves emitted by the jet into the surrounding space are surface waves; this corresponds to $m_2^2 > 0$.

(2) In the case of an incompressible jet, i.e., when $C_i = \infty$ ($i = 1, 2$) and $\varepsilon_i(\xi) = 0$, dispersion relation (1) becomes a quadratic equation in ξ. The requirement that the discriminant be nonnegative yields the following condition for the stability of the jet [3, 4]:

$$a^2 \leq \frac{[\nu G(\delta) + 1][\nu G(\delta)b_1^2 + b_2^2]}{\nu G(\delta)}. \tag{11}$$

In condition (11), the equality sign corresponds to the square of the critical velocity, a_{cr}^2, above which the instability develops at a given perturbation wavelength. The function $G(\delta)$, which is determined from formulas (4) with $m_1 = m_2 = 1$, is plotted in Fig. 1. We can see that, for a plane jet, this function increases monotonically from 0 to 1 whereas for a cylindrical jet, it decreases monotonically from ∞ to 1. Consequently, the geometry of the jet is important from the standpoint of its stability only in the case of long-wavelength perturbations such that $\delta < 1$.

In [4], condition (11) was analyzed in order to determine the minimum value of the critical velocity a_{cr} below which the jet is stable for any δ values. The results of this analysis are illustrated in Fig. 2, which

implies that, in the case at hand, the effect of the jet geometry is determined by the parameter

$$\beta = \frac{b_2}{\nu b_1} = \left(\frac{\rho_2 H_2^2}{\rho_1 H_1^2}\right)^{1/2}. \tag{12}$$

In the range $G(\delta) \leq 1$, corresponding to a plane jet, the following regular features can be revealed. If $\beta > 1$, then, as $G(\delta)$ increases from 0 to 1 (i.e., as δ increases from 0 to ∞), the squared critical velocity a_{cr}^2 decreases monotonically from ∞ to its minimum value at the point $G(\delta) = 1$:

$$a_{cr}^2\big|_{G(\delta) = 1} = (b_1 + b_2)^2 + \nu b_1^2(1 - \beta^2). \tag{13}$$

If $\beta < 1$, then the squared critical velocity a_{cr}^2 is a nonmonotonic function of $G(\delta)$ and its minimum value is reached at the extreme point $G(\delta) = \beta$ and is equal to

$$a_{cr}^2\big|_{min} = (b_1 + b_2)^2. \tag{14}$$

The situation with a cylindrical jet, corresponding to the range $G(\delta) \geq 1$, is opposite to that with a plane jet. Hence, for $\beta > 1$, a plane jet is more stable than a cylindrical jet, and vice versa for $\beta < 1$.

(3) Now, we consider the same problem but take into account a low compressibility such that $\mu_i \gg 1$ and $\mu_i > b_i$. In this case, we have $|\varepsilon_i(\xi)| \ll 1$, i.e., $m_i^2 > 0$ ($i = 1, 2$); therefore, the right-hand side of dispersion relation (1) is negative. This indicates that the necessary condition for the stability of the jet, i.e., the condition for the roots of dispersion relation (1) to be real, can be satisfied only when $N_1(\xi)$ and $N_2(\xi)$ have opposite signs. Under the conditions of a low compressibility such that $\mu_i > b_i$, expressions (5) and (6) imply that the inequality $m_i^2 > 0$ is satisfied when $N_i(\xi)$ and $D_i(\xi)$ have the same signs (here, D_i is the denominator of m_i^2), i.e., when they are both positive or both negative, which corresponds to the propagation of a transverse Alfvén wave or a longitudinal magnetosonic wave, respectively. Hence, the jet is stable when the hydrodynamic waves propagating inside and outside it are different in nature; moreover, a transverse Alfvén wave is generated on the side of the tangential discontinuity where the Alfvén speed is higher, and a longitudinal magnetosonic wave is generated on the side where the Alfvén speed is lower [5].

We assume that $G(\delta, \xi)$ is a slowly varying function of ξ and write dispersion relation (1) in the form

$$[\nu G(\delta, \xi^0) + 1]\xi^2 - 2a\xi + a^2$$
$$- \nu G(\delta, \xi^0)b_1^2 - b_2^2 = 0, \tag{15}$$

where ξ^0 is the root of dispersion relation (1) in the case of an incompressible jet, when the discriminant of the relation is zero at the critical jet velocity.

We expand the function $G(\delta, \xi^0)$ in powers of the small parameters $\varepsilon_i^0 = \varepsilon_i(\xi^0)$ and retain terms up to the first order to obtain

$$G(\delta, \xi^0) = G(\delta)(1 + \varepsilon), \qquad (16)$$

where

$$\varepsilon = \begin{cases} \dfrac{1}{2}\left[\varepsilon_2^0 - \varepsilon_1^0\left(1 - \dfrac{2\delta}{\sinh 2\delta}\right)\right] \text{ for a plane jet,} \\[2mm] \varepsilon_2^0\left[1 + \dfrac{\delta}{2}\left(\dfrac{K_1(\delta)}{K_0(\delta)} - \dfrac{K_0(\delta)}{K_1(\delta)}\right)\right] \\[2mm] -\varepsilon_1^0\left[1 + \dfrac{\delta}{2}\left(\dfrac{I_0(\delta)}{I_1(\delta)} - \dfrac{I_1(\delta)}{I_0(\delta)}\right)\right] \text{ for a cylindrical jet.} \end{cases} \qquad (17)$$

It can be shown that all the coefficients by ε_i^0 are positive; hence, we have

$$\varepsilon \begin{cases} >0, \quad \text{for} \quad \varepsilon_2^0 > 0, \quad \varepsilon_1^0 < 0, \\ <0, \quad \text{for} \quad \varepsilon_2^0 < 0, \quad \varepsilon_1^0 > 0. \end{cases} \qquad (18)$$

Comparing expressions (2) and (3) with expressions (7) and (8), we easily find that, for $\mu_i > b_i$, the quantities ε_i^0 and $N_i^0 = N_i(\xi^0)$ have the same signs.

Under the conditions of a low compressibility, dispersion relation (15) yields the following expression for the critical velocity:

$$a_{cr}^2 = \frac{[\nu G(\delta, \xi^0) + 1][\nu G(\delta, \xi^0) b_1^2 + b_2^2]}{\nu G(\delta, \xi^0)}. \qquad (19)$$

Since the right-hand side of expression (19) formally coincides with that of condition (11), we can analyze this expression in the above manner to obtain an analogous result. The only difference is that the extreme point, which is defined by the formula

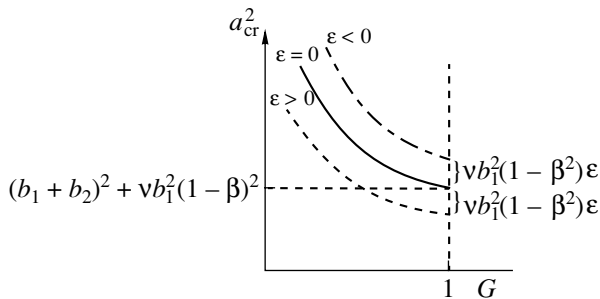$$G(\delta) = \frac{\beta}{1 + \varepsilon}, \qquad (20)$$

will be somewhat shifted to the right or to the left, depending on the sign of ε. Accordingly, we will arrive at another value of a_{cr}^2 at the point $G(\delta) = 1$:

$$a_{cr}^2\big|_{G(\delta) = 1} = (b_1^2 + b_2^2) + \nu b_1^2(1 - \beta^2) + \nu b_1^2(1 - \beta^2)\varepsilon. \qquad (21)$$

Let us now consider the difference between the squared critical velocities given by expression (19) and condition (11):

$$\Delta a_{cr}^2 = \frac{\nu b_1^2[G^2(\delta) - \beta^2]}{G(\delta)}\varepsilon. \qquad (22)$$

It is clear that the effect of compressibility on the stability of a plasma jet is governed by the sign of this difference: when $\Delta a_{cr}^2 > 0$, the compressibility stabilizes the jet, and, when $\Delta a_{cr}^2 < 0$, the effect is destabilizing. Let us analyze expression (22) with allowance for the results obtained in [4, 5].

Since the case in which $\beta > 1$ and $G(\delta) \leq 1$ is most optimal for the stability of a plane jet, the difference in the square brackets is negative. On the other hand, the inequality $\beta = (\rho_2 H_2^2/\rho_1 H_1^2)^{1/2} > 1$ can be satisfied at the expense of a jump in the magnetic field ($H_2 > H_1$, $\rho_2 \leq \rho_1$) or in the density ($H_2 \leq H_1$, $\rho_2 < \rho_1$). In the first case, we have $V_{A2} > V_{A1}$ and, as was mentioned above, $N_2^0 > 0$ ($\varepsilon_2^0 > 0$) and $N_1^0 < 0$ ($\varepsilon_1^0 < 0$). In this case, relationships (18) yield $\varepsilon > 0$ and $\Delta a_{cr}^2 < 0$, which indicates that the compressibility destabilizes the jet; moreover, the waves generated inside and outside a stable jet area were a magnetosonic wave and an Alfvén wave, respectively. In the second case, we have $V_{A2} < V_{A1}$ and, consequently, $N_2^0 < 0$ ($\varepsilon_2^0 < 0$), $N_1^0 > 0$ ($\varepsilon_1^0 > 0$), $\varepsilon < 0$, and $\Delta a_{cr}^2 > 0$. This indicates that the compressibility stabilizes the jet; in this case, the waves generated inside and outside the jet are an Alfvén wave and a magnetosonic wave, respectively.
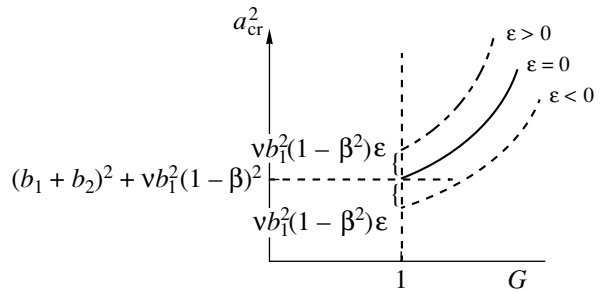

Fig. 3. Dependence of a_{cr}^2 on $G(\delta)$ for a plane compressible jet at $\beta > 1$.
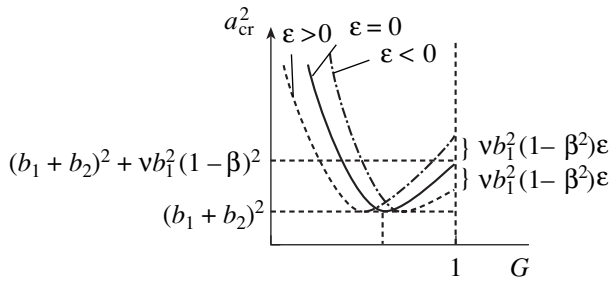

Fig. 4. Dependence of a_{cr}^2 on $G(\delta)$ for a cylindrical compressible jet at $\beta < 1$.

Fig. 5. Dependence of a_{cr}^2 on $G(\delta)$ for a compressible plane jet at $\beta < 1$.
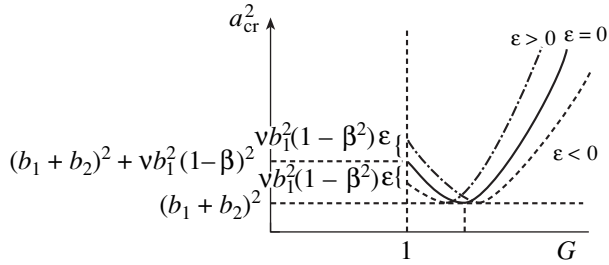

Fig. 6. Dependence of a_{cr}^2 on $G(\delta)$ for a compressible cylindrical jet at $\beta > 1$.

The optimal conditions for the stability of a cylindrical jet are $\beta < 1$ and $G(\delta) \geq 1$; accordingly, the difference in the square brackets of expression (22) is positive, i.e., the sign of Δa_{cr}^2 is determined by the sign of ε. By analogy with the previous case, we can show that, if $H_2 < H_1$ and $\rho_2 \geq \rho_1$, then we have $V_{A1} > V_{A2}$ and $\Delta a_{cr}^2 < 0$ by virtue of $\varepsilon < 0$. This indicates that the compressibility has a destabilizing effect; moreover, the

waves generated inside and outside the jet are an Alfvén wave and a magnetosonic wave, respectively. The results of this analysis are illustrated schematically in Figs. 3 and 4.

It is also of interest to consider conditions that are nonoptimal for the jet stability, namely, the condition $\beta < 1$ for a plane jet and the condition $\beta > 1$ for a cylindrical jet. The results of the corresponding analysis of expression (22) are demonstrated in Figs. 5 and 6. It can be seen that, under these conditions, the effect of compressibility on the jet stability depends not only on the relationship between the MHD parameters of the jet and the surrounding medium but also on the perturbation wavelength.

Finally, we can draw the following general conclusions. For a plasma jet of a certain (plane or cylindrical) geometry, the compressibility (which undoubtedly destabilizes the tangential discontinuity in the classical sense) can play a destabilizing or a stabilizing role, depending on the jet-flow conditions and the perturbation wavelength. In addition, the plasma jet operates as a generator of MHD waves, whose nature is also determined by the jet-flow conditions.

REFERENCES

1. V. G. Kirtskhalia, Soobshch. Akad. Nauk Gruz. **117**, 282 (1983).
2. A. I. Gvelesiani, G. V. Dzhandieri, and V. G. Kirtskhalia, Soobshch. Akad. Nauk Gruz. **121**, 61 (1986).
3. E. N. Parker, Astrophys. J. **39**, 690 (1964).
4. V. G. Kirtskhalia, I. A. Zhvania, and A. I. Gvelesiani, Byull. Akad. Nauk Gruz. **161** (31), 48 (2000).
5. V. G. Kirtskhalia and A. A. Rukhadze, Kratk. Soobshch. Fiz. **54** (11), 50 (2003).

Translated by O. Khadin

BRIEF COMMUNICATIONS

Optical and Electrical Properties of Propolis Films

S. I. Drapak*, I. T. Drapak**, and Z. D. Kovalyuk*

* Institute of Problems of Materials Science, Chernovtsy Division,
National Academy of Sciences of Ukraine, Chernovtsy, 58000 Ukraine
e-mail: chimsp@unicom.cv.ua
** Fed'kovich National University, Chernovtsy, 58012 Ukraine
Received March 19, 2004

Abstract—The basic electrical and optical properties of films made of propolis, which is a natural, biologically active, organic compound, are studied. Photoluminescence is found to exist at room temperature with a maximum at 434 nm. The conduction activation energy in the temperature range 283–300 K is 2.9 eV and correlates with the optical band gap. © *2004 MAIK "Nauka/Interperiodica"*.

The application of different-type organic semiconducting materials in electronics is widening continuously [1, 2]. This circumstance stimulates studies of the physical properties of both synthesized and biological (skin, green leaves, albumen, etc.) materials and their possible application as components of optoelectronic devices [3–5]. It was shown in [6–8] that semiconductor (p-InSe, n-Si)–propolis hybrid photosensitive structures, where the natural organic material behaves as an p-type semiconductor, can be fabricated.

Propolis (bee glue), which is extensively used in pharmacology, perfumery, and cosmetology, is a product of the life of honeybees and is a very complex mixture of organic materials; generally, it consists of tarry matter (50–55%), wax (up to 30%), and ethereal oils and balsams (about 10%). Moreover, propolis (hereafter, P) contains organic acids, antibiotics, a number of vitamins and microelements (Al, V, Fe, Ca, Si, Mn, and Sr), and some natural enzymes (e.g., carotene) [9]. The methods of medical treatment with P and its chemical composition were analyzed in numerous works (e.g., see [10]). Every new study of the chemical composition still reveals new components of this compound. When studying the physical properties of P, researchers mainly determine its density, the temperature ranges of its aggregation states, and so on.

In this work, we determined for the first time the basic electrical and optical parameters of films made of this natural, biologically active, organic compound. We recorded its transmission spectrum in the wavelength range $\lambda = 350$–600 nm and its photoluminescence spectrum, determined the electrical conductivity at room temperature, and found the temperature dependence of the conductivity.

To prepare films of uniform thickness, a drop of a 10% alcohol solution of P was applied onto a sapphire substrate fixed on a centrifuge; it rotated until the alcohol completely evaporated (no longer than 1 min). The thickness of the P films was 15–30 μm. As a solvent, we used 96% ethyl alcohol. The conductivity of the films was measured by the standard two-probe technique [11]. Silver was used for electrical contacts.

Figure 1 shows the optical properties of the P films. As can be seen, the P films are transparent in the wavelength range from 410 to 600 nm ($T \approx 90\%$ at film thickness $d \approx 20$ μm) (Fig. 1, curve 1); the long-wavelength tail of the optical absorption edge in P is strongly diffused; and the absorption coefficient α changes by an order of magnitude (from 10^3 to 10^2 cm^{-1}) in the wavelength range from 380 to 410 nm (Fig. 1, curve 2), which corresponds to photon energies of 3.26–3.02 eV.

Luminescence of P is observed when it is excited by light with a wavelength of 337 nm. The photoluminescence spectrum is shown in Fig. 1 (curve 3). The maximum in the spectral dependence corresponds to 434 nm (2.857 eV). Noticeable inflection points on either side of the maximum indicate the presence of several types of luminescence centers in P, which is characteristic of both semiconductors having different impurities and complex molecular mixtures [12].
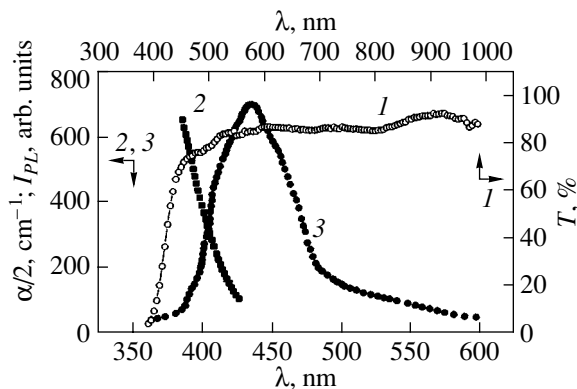

Fig. 1. (1) Transmission spectrum, (2) optical absorption coefficient, and (3) photoluminescence spectrum of the propolis films at $T = 288$ K.
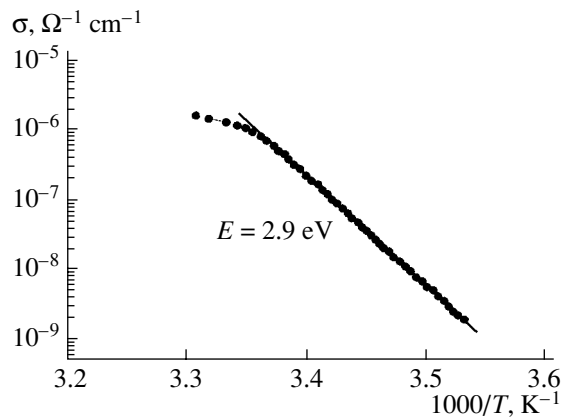
Fig. 2. Temperature dependence of the conductivity of the propolis films.

Figure 2 shows the temperature dependence of the conductivity $\sigma(T)$ of P. As can be seen, the conductivity has a constant activation energy (≈ 2.9 eV) in the temperature range 283–300 K. As the temperature increases further, the growth in the conductivity decelerates significantly, which can be caused by a number of factors. First, the deceleration of the increase in the conductivity with increasing temperature is characteristic of materials with a high content of impurities and can be related to a decrease in the potential barrier heights [13]. Second, this behavior of $\sigma(T)$ is typical of noncrystalline semiconductors when all charge carriers have passed from localized centers to the valence or conduction band under the action of temperature [14]. Finally, if we assume that the P films are somewhat ordered at $T < 300$ K (by analogy with inorganic semiconductors, which have a long-range order), an increase in temperature results in disordering of the P structure and the formation of localized states in the energy structure of this material (at 30°C, the wax entering into the P composition begins to soften). Of course, each assumption requires additional studies.

Thus, the results of the first studies of the conductivity and optical properties of propolis indicate that this natural, biologically active material is a semiconductor with an optical band gap of ≈ 3 eV, which can be used to design optoelectronic devices of different types. Moreover, P films have some advantages over other artificially synthesized commercial organic materials that are used as components in electronic devices [15]. Specifically, these films have a substantially lower resistivity (by four to five orders of magnitude) and a wider optical band gap. Another advantage of P over its analogs studied in the literature (skin, green leaves, albumen, etc.) [4, 5] and some organic semiconductors is its high environmental resistance [9].

REFERENCES

1. R. Rinaldi, E. Branca, R. Cingolani, *et al.*, Appl. Phys. Lett. **78**, 3541 (2001).

2. T. A. Yourre, L. I. Rudaya, N. V. Klimova, and V. V. Shamanin, Fiz. Tekh. Poluprovodn. (St. Petersburg) **37** (7), 73 (2003) [Semiconductors **37**, 807 (2003)].

3. C. Okazaki, Jpn. J. Appl. Phys. **37**, Part 1, 983 (1998).

4. V. Yu. Rud', Yu. V. Rud', and V. Kh. Shpunt, Zh. Tekh. Fiz. **70** (2), 114 (2000) [Tech. Phys. **45**, 255 (2000)].

5. V. Yu. Rud', Yu. V. Rud', and V. Kh. Shpunt, Fiz. Tekh. Poluprovodn. (St. Petersburg) **31** (2), 129 (1997) [Semiconductors **31**, 97 (1997)].

6. S. I. Drapak, A. P. Bakhtinov, V. B. Orletskiĭ, and Z. D. Kovalyuk, in *Proceedings of the 1st International Meeting on Applied Physics "APHYS-2003," Badajoz, 2003.*

7. S. I. Drapak, V. B. Orletskiĭ, Z. D. Kovalyuk, and V. V. Netyaga, Pis'ma Zh. Tekh. Fiz. **29** (20), 69 (2003) [Tech. Phys. Lett. **29**, 867 (2003)].

8. S. I. Drapak, V. B. Orletskiĭ, V. D. Fotiĭ, and Z. D. Kovalyuk, Prikl. Fiz. (2004) (in press).

9. V. V. Dontsov and V. I. Dontsov, *Medicinal Plants and Apiculture Products* (Floks, Nizhni Novgorod, 1992) [in Russian].

10. N. G. Takaisi-Kikuni, Planta Med. **60**, 222 (1994).

11. N. Ya. Karkhanina, in *Semiconductor Technology* (GITL, Kiev, 1961) [in Russian].

12. C. A. Parker, *Photoluminescence of Solutions* (Elsevier, Amsterdam, 1968).

13. P. T. Oreshkin, *Physics of Semiconductors and Dielectrics* (Vysshaya Shkola, Moscow, 1977) [in Russian].

14. N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford, 1979; Mir, Moscow, 1982).

15. E. A. Lebedev, M. Ya. Goĭkhman, M. E. Kopman, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **37**, 844 (2003) [Semiconductors **37**, 816 (2003)].

Translated by K. Shakhlevich

BRIEF
COMMUNICATIONS

Nonmonotonic Absorption in Yttrium–Barium Cuprates

B. M. Gorelov and V. A. Sidorchuk

Institute of Surface Chemistry, National Academy of Sciences of Ukraine, Kiev, 03164 Ukraine
e-mail: user@surfchem.freenet.kiev.ua
Received March 30, 2004

Abstract—It is shown that the adsorption, solubility, and diffusion coefficient of water molecules in metal-oxide cuprate $YBa_2Cu_3O_{7-\delta}$ are nonmonotonic functions of the oxygen content. The behavior of these quantities is determined by the superposition of two processes: the change in the charge state of atoms in intermediate layers and the electron density distribution between cuprate and intermediate layers upon filling of O1 positions by oxygen. © 2004 MAIK "Nauka/Interperiodica".

It is well known that water molecules are absorbed by the crystal lattice of metal-oxide cuprates $YBa_2Cu_3O_{7-\delta}$ [1]. If adsorption takes place at room temperature and under saturated vapor pressure, a layer of physically bound water is formed; H_2O molecules from this layer are implanted into the lattice, where they form four localized states at interstices of the BaO and Cu1O intermediate layers. The filling of these states increases with time [2]. It can be expected that binding of molecules at the surface and their diffusion to the bulk of the crystal depend on the charge states of atoms in the lattice, which determine the energy barriers of chemisorption and diffusion of molecules. Since the charge states of atoms are determined by the oxygen content in the system [3], water absorption may be a function of the number of oxygen atoms in the lattice.

Here, we analyze the dependence of the absorption of water molecules on the content of O1 atoms in the system $YBa_2Cu_3O_{7-\delta}$. We studied disperse samples of $YBa_2Cu_3O_{7-\delta}$ with an oxygen content varying in the range $0 \le \delta < 1$. Particles had the shape of flakes ~20 μm in diameter and ~10 μm in thickness. Adsorption took place at room temperature in an atmosphere of saturated vapor (a vapor pressure of 18.7 Torr) on samples preliminarily annealed in a vacuum of ~10^{-3} Torr at $T < 150°C$. We used doubly distilled water as an adsorbate. Adsorption was measured with the help of a McBain balance. The specific surface was determined on the GKh-1 setup from nitrogen desorption.

Figure 1 illustrates the adsorption kinetics of water molecules in $YBa_2Cu_3O_{7-\delta}$ samples upon a variation of δ in the interval $0 \le \delta < 1$. After the formation of a layer of physically bound water at the surface of particles, the adsorption coefficient a smoothly increases with time t during $t \le 90$ min and exhibits a tendency to saturation at $t > 2000$ min. The slow increase in the adsorption coefficient is determined by the implantation and diffusion of H_2O molecules in the crystal lattice. A peculiar feature of adsorption upon an increase in δ is a non-

monotonic variation of a in the interval $0.1 \le \delta \le 0.4$ and its sharp increase for $\delta > 0.4$, which is clearly manifested in the behavior of limiting adsorption and solubility (Fig. 2a). The dependences $a_\infty(\delta)$ and $N_0(\delta)$ are characterized by a nonmonotonic decrease in a_∞ and N_0 with increasing δ in the interval $0.1 \le \delta \le 0.35$ with a minimum at $\delta \cong 0.25$, a sharp increase for $\delta > 0.35$, and a slight increase for $\delta > 0.45$.

The behavior of the effective diffusion coefficient D_e of water molecules, derived from the expression [4]

$$\frac{a(t)}{a_\infty} = \frac{2S_1}{V}\sqrt{\frac{D_e t}{\pi}}, \tag{1}$$

where V is the volume and $S_1 = \rho V s$ is the area of the outer surface of particles (ρ is the density and s is the specific surface), depending on the oxygen content is
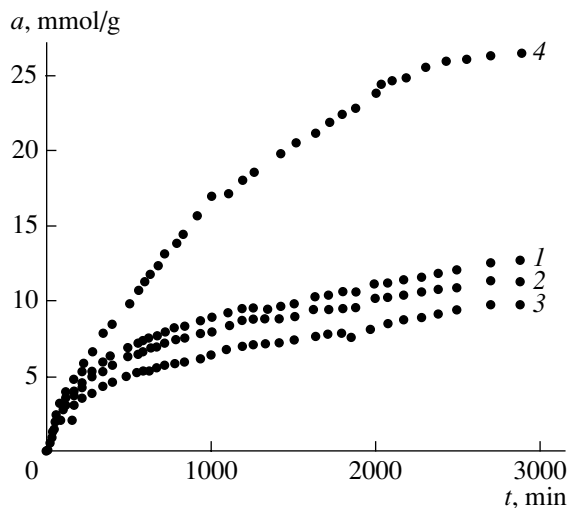

Fig. 1. Kinetics of adsorption of water molecules in $YBa_2Cu_3O_{7-\delta}$ with $\delta = 0$ (*1*), 0.2 (*2*), 0.24 (*3*), and 0.44 (*4*).
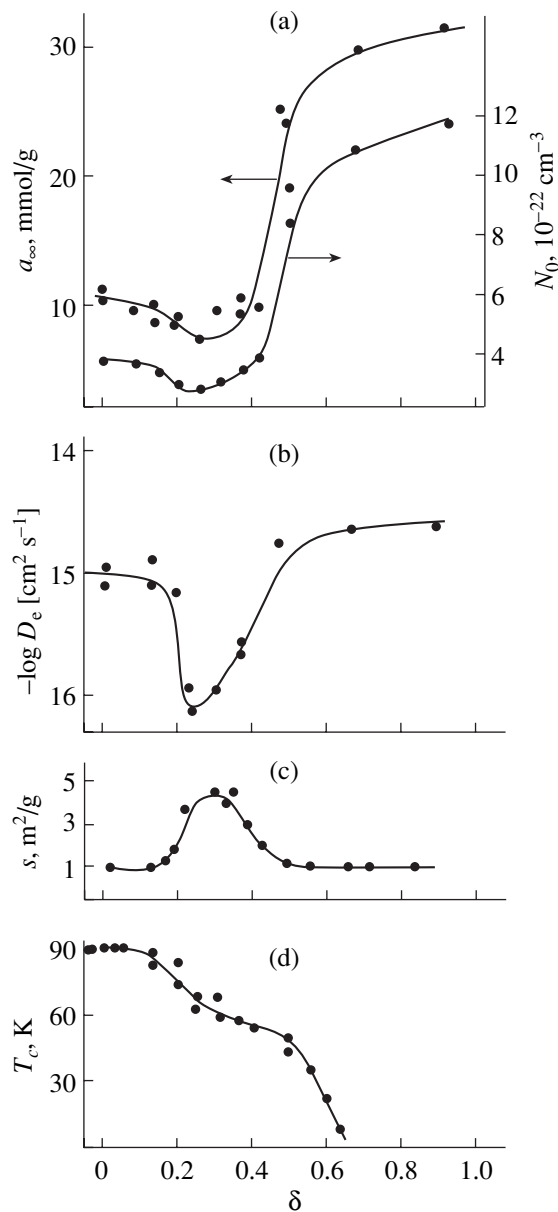
Fig. 2. Dependences of the limiting absorption and solubility (a), the effective diffusion coefficient of water (b), the specific surface (c), and the superconducting transition temperature (borrowed from [5]) (d) on the oxygen content in $YBa_2Cu_3O_{7-\delta}$.

similar to the behavior of function $N_0(\delta)$ (Fig. 2b). In determining D_e, we assumed that particles of the samples are of the same size and that the rate of transition of an H_2O molecule from the free to the bound state at the surface of the particles is much higher than the diffusion rate in the crystal lattice.

It should be noted that the specific surface increases nonmonotonically in the interval of the nonmonotonic decrease in a_∞, N_0, and D_e (Fig. 2c). In addition, the region of nonmonotonic variation of U and D_e for $0.1 < \delta \le 0.6$ coincides with the interval in which the super-

conducting transition temperature T_c changes stepwise, while the range of $\delta > 0.6$, where the values of U and D_e change insignificantly, corresponds to the interval of oxygen content in which superconductivity is not observed (Fig. 2d) [5].

Thus, the solubility and the diffusion coefficient of water in $YBa_2Cu_3O_{7-\delta}$ are nonmonotonic functions of the number of O1 atoms in intermediate layers. Since the solubility is proportional to the number of vacancies in the lattice and the binding energy of water atoms with the atomic surroundings,

$$U = \frac{1}{8\pi\varepsilon\varepsilon_0}\sum_{i,\,\alpha} q_\alpha^0 V_\alpha^i \qquad (2)$$

(where ε_0 is the permittivity, q_α^0 is the charge state of cuprate atoms of sort α, and V_α^i is the electrostatic potential produced by jth atoms of βth molecules of water at an ith atom), a decrease in N_0 upon a smooth increase in the number of O1 vacancies in the lattice with increasing δ indicates that H_2O molecules do not fill O1 vacancies and the behavior of N_0 can be attributed to the nonmonotonic variation of U and the charge state of atoms in the lattice. As the number of O1 atoms grows ($\delta \longrightarrow 0$), the binding energy decreases first insignificantly for $\delta > 0.6$ and then sharply in the interval $0.25 \le \delta \le 0.6$, after which it smoothly increases for $0.1 \le \delta \le 0.25$ and virtually does not change for $\delta < 0.1$.

In addition, the diffusion coefficient in a multicomponent system is given by [6]

$$D = ukT\left(1 + \frac{\partial \ln\gamma}{\partial \ln N}\right), \qquad (3)$$

where u, γ, and N are the mobility, the activity, and the number of H_2O molecules. The mobility is connected with diffusion rate v and chemical potential μ by the expression

$$v = -\frac{u}{N_A}\frac{\partial\mu}{\partial x}; \qquad (4)$$

consequently, the behavior of $D(\delta) = D_e a_\infty/c_0$ (c_0 is the concentration of molecules at the particle surface) is also determined by the nonmonotonic variation of $\partial\mu/\partial x$ upon a smooth decrease in δ. On the other hand, $D = D_0\exp(-E/kT)$, where k is the Boltzmann constant and the activation energy $E \sim U$ [6]; in this case, the behavior of $D(\delta)$ for $\delta \longrightarrow 0$ is determined by the nonmonotonic decrease in $E(\delta)$. It is important that the behavior of U, E, and $\partial\mu/\partial x$ for $\delta \longrightarrow 0$ can be associated with nonmonotonic variation of the charge state of lattice atoms.

The increase in the number of O1 atoms in $YBa_2Cu_3O_{7-\delta}$ for $\delta \longrightarrow 0$ changes the charge state of the atoms constituting the intermediate layers. The charge state of Ba and Cu1 decreases from 2.08 and

1.46 to 1.45 and 0.81, respectively; the charge of the O4 atom decreases from –2.08 to –0.67, and the number of O1 atoms with a charge of –1.76 increases [4]. In addition, the lattice parameter c decreases, but the apex atom O4 moves away from Cu1 and approaches Cu2 [7]. In this case, holes are redistributed between the $d_{x^2-y^2}$ and d_{z^2} orbitals of the Cu2 atom when holes are localized at $d_{x^2-y^2}$, and the electron density becomes extended along the **c** axis [4]. Therefore, the nonmonotonic behavior of U, E, and $\partial\mu/\partial x$ is probably determined by two processes that affect the value of $\partial\mu/\partial x$, viz., a smooth decrease in the charge state of Ba, Cu1, and O4 upon an increase in the number of O1 atoms, leading to a decrease in U and E, and a redistribution of the electron density in Cu2 orbitals upon a convergence with the O4 atom in the interval $0.1 \leq \delta \leq 0.25$, which might result in an increase in V_α^i and, accordingly, in the values of U and E.

It should be noted that the electron density redistribution matches the increase in T_c in the interval $0.1 \leq \delta \leq 0.25$ for $\delta \longrightarrow 0$, since it leads to an increase in the number of holes, p, in cuprate layers and in the conduction band formed by the orbitals $d_{x^2-y^2}p_x(p_y)$ of Cu2 and O2 (O3) atoms and in temperature $T_c(p) = T_{cm}[1 - 82.6(p - 0.16)^2]$, where T_{cm} is the maximal superconducting transition temperature in $YBa_2Cu_3O_{7-\delta}$ [8].

Thus, the nonmonotonic absorption of water molecules in $YBa_2Cu_3O_{7-\delta}$ is due to changes in the charge state of the atoms constituting intermediate layers and the electron density distribution between cuprate and intermediate layers upon a variation of the oxygen content.

REFERENCES

1. L. L. Makarshin, D. V. Andreev, and V. N. Paramonov, Usp. Khim. **69**, 307 (2000).

2. B. M. Gorelov, D. V. Morozovskaya, V. M. Pashkov, and V. A. Sidorchuk, Zh. Tekh. Fiz. **70** (9), 50 (2000) [Tech. Phys. **45**, 1147 (2000)].

3. V. E. Gusanov, Fiz. Nizk. Temp. **21**, 805 (1995).

4. D. P. Timofeev, *Kinetics of Adsorption* (Izd. Akad. Nauk SSSR, Moscow, 1962) [in Russian].

5. *Chemistry of High-Temperature Superconductors,* Ed. by D. L. Nelson, M. S. Whittingham, and T. F. George (Am. Chem. Soc., Washington, 1987; Mir, Moscow, 1988).

6. B. I. Boltaks, *Diffusion in Semiconductors* (Fizmatgiz, Moscow, 1961; Academic, New York, 1963).

7. R. C. Baetzold, Phys. Rev. B **42** (1A), 56 (1990).

8. G. V. M. Williams, J. L. Tallon, R. Michalak, and R. Dupree, Phys. Rev. B **54**, 6909 (1996).

Translated by N. Wadhwa

BRIEF
COMMUNICATIONS

Effect of a SrF$_2$ Insulating Film on the Luminescent Properties of n-InP

V. V. Agaev, V. A. Sozaev, and G. I. Yablochkina

North-Caucasian Institute of Mining and Metallurgy (State Technological University),
Vladikavkaz, 362021 Russia

e-mail: skgtu@skgtu.ua

Received March 30, 2004

Abstract—It is shown that SrF$_2$ can be used as an effective protective coating for indium phosphide films, since it decreases the surface recombination rate of the films due to lattice matching between SrF$_2$ and InP. As a result, the external quantum yield of radiative recombination increases. © *2004 MAIK "Nauka/Interperiodica".*

The recording and numerical estimation of the intensity of UV radiation have recently became a key issue in view of ecological monitoring and wide application of UV sources in science, engineering, medicine, and industry. These factors require the diversification of photodetector-based functional measuring instruments [1, 2].

Along with GaP-, Si-, Ge, and InGaAsP-based photodetectors [3], photodetectors with the upper layer made of n-InP are promising for the UV region. In such photodetectors, the p–n junction should be located as close as possible to the surface; in this case, however, surface recombination may substantially affect the photodetector efficiency.

Because of the high internal quantum yield (90%) and a low surface recombination rate in n-InP, the role of the effects of multiple passage and reradiation, which are important for optoelectronic devices, becomes more significant [4]. These effects provide effective excitation transfer in n-InP [5].

The surface recombination rate in n-InP is only 2×10^4 cm/s (for comparison, this rate in n-GaAs is 3×10^5 cm/s [6]), which makes it possible to design a high-sensitivity UV photodetector. However, the open surface of n-InP is unstable in time [7]. The InP surface can be stabilized with protective coatings. There are many insulating coatings, such as native oxide In(PO$_4$)$_2$ and oxides of other compounds (Al$_2$O$_3$, SiO$_2$, etc.), that can be used to protect n-InP. Interest in fluoride-based films has quickened in the past few years [8, 9]; for InP, such films are made of SrF$_2$.

A SrF$_2$ film ensures transparency in the near-UV region and increases the surface recombination rate, since the lattice parameter of SrF$_2$ is close to that of InP. A SrF$_2$ film forms a heterojunction at the interface with InP, which not only effectively protects the surface but also reduces the surface recombination in InP (which is evidenced by the increase in the external quantum yield of radiative recombination).

The epitaxial InP films were produced by liquid-phase epitaxy [4]. The epitaxial-layer thickness was 2–3 μm, and the concentration of uncontrolled impurities was $n = 5 \times 10^{17}$ cm^{-3}, as determined from metal–insulator breakdown.

Each as-grown film was divided into three parts. One part was not treated (the reference part), and the two others were coated with SrF$_2$, CaF$_2$, or SiO + SiO$_2$ insulating films.

The insulating films were fabricated by vacuum deposition; their thickness was varied from 0.4 to 1 μm.

To study the photoluminescence properties, we used a setup similar to that described in [4]. A continuous 1-W Ar laser with wavelength $\lambda = 0.514$ μm was used as an excitation source.

The spectral dependences of all three parts of the epitaxial film are shown in the figure. To estimate the
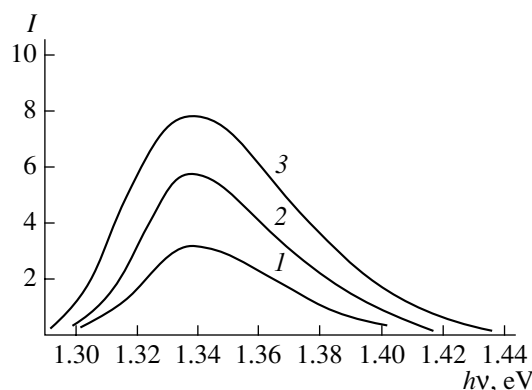
Luminescence spectra of n-InP with and without a insulating film: (*1*) the mirror surface of the epitaxial n-InP film, (*2*) the mirror surface of the epitaxial film coated by a SiO + SiO$_2$ insulating film, and (*3*) the mirror surface of the epitaxial film coated by a SrF$_2$ insulating film.

external quantum yield, the spectra of all the samples were compared with the spectrum of a calibrated heterostructure. The external quantum yield of the radiative recombination in the InP epitaxial films was estimated under the same experimental geometry and at a constant quantum sensitivity of the photodetector.

The external quantum luminescent efficiency was $\eta_e = 2.2\%$ for the pure (uncoated) sample, $\eta_e = 4\%$ for the film with the SiO + SiO$_2$ coating, and $\eta_e = 6.1\%$ for the film coated with SrF$_2$.

To compare the experimental data with theory, we calculated the external quantum yield of luminescence for the uncoated and coated InP epitaxial films. We assumed that the internal quantum yield of luminescence is 100% in a narrow excited region and zero in the unexcited region, the reradiation effects are absent, the sample has a mirror surface, and the output emission is limited by the effects of total internal reflection:

$$\eta_e = \frac{2n_0^3}{n(n_0 + n)^2}.$$

Our calculations show that the external quantum yield is $\eta_e = 2.6\%$ for the uncoated n-InP epitaxial films and does not exceed 4.5% for the samples coated with the insulating films ($n_0 = 1.43$). These calculations agree well with the results obtained for some n-InP films with the SiO + SiO$_2$ coating. An increase in the external quantum yield of the films with the SrF$_2$ coating by a factor of more than three as compared to the uncoated films is thought to be caused by a decrease in the interface recombination rate, since SrF$_2$ has a cubic lattice with a parameter of 5.79 Å, which virtually coincides with the InP lattice parameter. This fact results in a significant decrease in the nonradiative recombina-tion at the InP–SrF$_2$ interface and, hence, in an increase in the reradiation, which is evidenced by increasing the line halfwidth and a substantial increase in the external quantum yield (this value exceeds the data calculated using simple theoretical assumptions).

Thus, SrF$_2$ insulating films can serve as effective protective coatings for InP epitaxial films.

REFERENCES

1. V. I. Stafeev and I. D. Anisimov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **28**, 461 (1994) [Semiconductors **28**, 281 (1994)].
2. I. M. Anisimova, I. M. Vikulin, F. A. Zaitov, and Sh. D. Kurmashev, *Semiconductor Photodetectors: Ultraviolet, Visible, and Near-Infrared Spectral Regions* (Radio i Svyaz', Moscow, 1984) [in Russian].
3. A. M. Filachev, V. P. Ponomarenko, I. I. Taubkin, *et al.*, Prikl. Fiz., No. 6, 52 (2002).
4. D. Z. Garbuzov, V. V. Agaev, and A. T. Gorelenok, Fiz. Tekh. Poluprovodn. (Leningrad) **16**, 1538 (1982) [Sov. Phys. Semicond. **16**, 986 (1982)].
5. D. Z. Garbuzov, V. G. Agafonov, V. V. Agaev, *et al.*, Fiz. Tekh. Poluprovodn. (Leningrad) **17**, 2168 (1983) [Sov. Phys. Semicond. **17**, 1385 (1983)].
6. G. A. Hoffman and J. Gerritsen, J. Appl. Phys. **51**, 1603 (1980).
7. T. S. Kim, S. D. Lester, and B. G. Streetman, J. Appl. Phys. **61**, 2072 (1987).
8. A. G. Banshikov, A. V. Kimel', V. V. Pavlov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **42**, 884 (2000) [Phys. Solid State **42**, 909 (2000)].
9. S. V. Gastev, S. É. Ivanova, N. S. Sokolov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **44**, 1385 (2002) [Phys. Solid State **44**, 1445 (2002)].

Translated by K. Shakhlevich