Technical Physics, Vol. 50, No. 7, 2005, pp. 815–820. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 1–5.
Original Russian Text Copyright © 2005 by Korovin.

THEORETICAL AND MATHEMATICAL PHYSICS

Pressure Pulse Induced by a Pulsed Electric Current in a Cylindrical Liquid Conductor

V. M. Korovin
Institute of Mechanics, Moscow State University, Vorob'evy Gory, Moscow, 119192 Russia
e-mail: korovin@imec.msu.ru

Received October 11, 2004

Abstract—An analogue of z pinch is considered for a finite-conductivity liquid in a cylindrical nonconducting tube through which an exponentially decaying current pulse is passed. The distributions of magnetic induction, current density, and volume electromagnetic force pressure in the conducting liquid cylinder are found by solving an equation for a quasi-stationary electromagnetic field and the equation of magnetohydrostatics with the operational method. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

One method of monitoring the contamination of the metal melt by foreign inclusions (in particular, nonconducting solid particles) is based on passing current pulses through a quiescent liquid under test contained in a long cylindrical nonconducting tube [1]. The essence of such a method is based on the electromagnetic buoyancy force acting on foreign inclusions in a current-carrying liquid.

The electromagnetic buoyancy force arises when the pressure initial (in the absence of the current) distribution in the liquid changes under the action of a volume electromagnetic force, which results from interaction of the electric current with its self-magnetic field (Lorentz force). This force expels particles with a conductivity lower than that of the surrounding current-carrying liquid, while higher conductivity particles move inward the liquid [2]. Naturally, the gravitational field does not influence the magnitude and direction of the electromagnetic buoyancy force.

When the current density distribution in a liquid-metal cylinder is stationary and axisymmetric, the distributions of the electromagnetic field and Lorentz force pressure are described by an exact solution to the magnetohydrodynamic equations used in the z pinch theory [3–6].

For particles of a simple shape (spherical or cylindrical), the results of calculating the electromagnetic buoyancy force (EBF) by the expressions for the pressure produced by a uniform electric current passing through a liquid cylinder are given in a first approximation in [2, 7] (electromagnetic field distortions caused by the particles are neglected).

It is well known that a high-frequency current concentrates near the surface of a conductor (the skin effect) [3, 8]. The current density distribution over the conductor's cross section may also be nonuniform when a solitary current pulse is applied. For this reason, finding the nonstationary electromagnetic field distribution in a liquid-metal cylinder becomes a key issue in the EBF calculation.

In this work, we use the equations of quasi-stationary electromagnetic field and magnetohydrostatics to derive expressions for the induction, current density, and pressure in a liquid conducting cylinder with a current whose intensity (density integrated over the cylinder's cross sections) exponentially decays with time.

STATEMENT OF THE PROBLEM

Let us consider a long nonconducting cylindrical tube closed at both ends and completely filled with a quiescent homogeneous conducting liquid. At the zero time $t = 0$, an electric current is passed through the liquid along the axis of the cylinder. It is assumed that the current intensity (i.e., density integrated over the cross section of the liquid cylinder) varies by the law

$$g(t) = I\exp(-\alpha t)\cos\omega t, \quad I > 0, \quad \alpha > 0. \qquad (1)$$

We use the cylindrical frame of reference (r, φ, z) where the z axis is coincident with the axis of the tube and codirected with the current. Ignoring the effect of the end faces, though which the current is applied to the liquid, we assume that the desired functions, namely, magnetic induction $\mathbf{B} = (0, B, 0)$, current density $\mathbf{j} = (0, 0, j)$, and electric pressure p_e due to the current pulse, depend on variables r and t.

The system being considered represents an equilibrium magnetohydrodynamic configuration, z pinch [3–6], since the volume density of Lorentz force $\mathbf{j} \times \mathbf{B}$ is potential and has only the radial component whose action on the liquid is counterbalanced by the gradient of the nonstationary part of pressure $p_e(r, t)$. Clearly, the total pressure in the liquid, which is the sum of p_e and the hydrodynamic pressure, depends on the position of the tube relative to the vertical axis and, generally, is a function of r, φ, z, and t.

The distributions of the magnetic induction, current density, and the nonstationary part of the pressure in the liquid cylinder are described by the equations [3]

$$\frac{\partial B}{\partial t} = \nu\left(\frac{\partial^2 B}{\partial r^2} + \frac{1}{r}\frac{\partial B}{\partial r} - \frac{B}{r^2}\right), \tag{2}$$

$$j = \frac{1}{\mu_0 r}\frac{\partial (rB)}{\partial r}, \tag{3}$$

$$\frac{\partial p_{\mathrm{e}}}{\partial r} = -jB. \tag{4}$$

Here, $\nu = (\mu_0\sigma)^{-1}$ is the magnetic viscosity, $\mu_0 = 4\pi \times 10^{-7}$ N/m is the magnetic constant, and σ is the conductivity of the liquid. Having calculated the current density, we can determine electric field $\mathbf{E} = (0, 0, E)$ from the Ohm's law $j = \sigma E$.

At the zero time, the magnetic field inside the liquid cylinder is absent,

$$B(r, 0) = 0. \tag{5}$$

Let us find the boundary condition for the magnetic field at the surface of the liquid cylinder. For this purpose, we take advantage of the fact that the total current passing through the cylinder is known.

In view of (1), we find from differential form (3) of the Ampere law that

$$B(c, t) = \frac{\mu_0}{2\pi c} I \exp(-\alpha t)\cos\omega t, \tag{6}$$

where c is the radius of the liquid cylinder.

The problem of calculating the magnetic induction defined by (2), (5), and (6) refers to problem with discontinuous initial conditions [8]. This problem describes the diffusion of the magnetic field across the conducting cylinder in the case when an external electric current of intensity I is initially ($t = 0$) concentrated at the lateral ($r = c$) surface of the cylinder. Consequently, the tangential component of the magnetic field (the total field in our case) experiences a discontinuity at $r = c$, $t = 0$.

DISTRIBUTIONS OF THE ELECTROMAGNETIC FIELD AND PRESSURE

The boundary-value problem stated by (2) and (6) with initial condition (5) is solved by the operational method [9]. Let $\mathcal{L}$ be the operator of Laplace transformation (in time) and s be a parameter of this transformation. We put $\mathcal{L}[B(r, t)] = \mathcal{B}(s, r)$. Passing to transforms in induction equation (2) and boundary condition (6), we obtain, in view of initial condition (5),

$$\frac{d^2\mathcal{B}}{dr^2} + \frac{1}{r}\frac{d\mathcal{B}}{dr} - \left(\frac{s}{\nu} + \frac{1}{r^2}\right)\mathcal{B} = 0,$$

$$\mathcal{B}\big|_{r = c} = \frac{\mu_0 I}{2\pi c}\frac{s + \alpha}{(s + \alpha)^2 + \omega^2}.$$

Substitution $u = i\lambda r$ and $\lambda = \sqrt{s/\nu}$, where i is the imaginary unity, yields

$$\begin{gathered} \frac{d^2\mathcal{B}}{du^2} + \frac{1}{u}\frac{d^2\mathcal{B}}{du} + (u^2 - 1)\mathcal{B} = 0, \\ \mathcal{B}\big|_{u = i\lambda c} = \frac{\mu_0 I}{2\pi c}\frac{s + \alpha}{(s + \alpha)^2 + \omega^2}. \end{gathered} \tag{7}$$

Particular solutions of differential equation (7) are [9] the Bessel functions of the first, $J_1(u)$, and second, $Y_1(u)$, kind, respectively; however, only function $J_1(u)$ satisfies the boundedness condition at $u = 0$. Therefore, $\mathcal{B} = AJ_1(u)$, where A is an arbitrary constant. Boundary condition (7) leads to

$$A = \frac{\mu_0 I}{2\pi c}\frac{s + \alpha}{(s + \alpha)^2 + \omega^2}\frac{1}{J_1\left(i\sqrt{\frac{s}{\nu}}c\right)}.$$

Thus, the solution of operator problem (7) has the form

$$\mathcal{B}(s, u) = \frac{\mu_0 I}{2\pi c}\frac{s + \alpha}{(s + \alpha)^2 + \omega^2}\frac{J_1\left(i\sqrt{\frac{s}{\nu}}r\right)}{J_1\left(i\sqrt{\frac{s}{\nu}}c\right)}. \tag{8}$$

Function $\mathcal{B}(s, r)$ defined by equality (8) is a single-valued function in s. In accordance with the expansion theorem following from the theory of inverse Laplace transformation [9], original $B(r, t)$ can be written as a sum of the residues of function $\mathcal{B}(s, r)\exp(st)$ calculated in all the poles of function $\mathcal{B}(s, r)$. It is easy to see that function $\mathcal{B}(s, r)$ has an infinite number of first-order poles. In the complex domain, the poles of s are points $s_{01} = -\alpha + i\omega$ and $s_{02} = -\alpha - i\omega$, as well as points $s_k = -\nu\beta_k^2/c^2$ ($k = 1, 2, 3, \ldots$) defined by roots β_k of the equation $J_1(u) = 0$ (point $s = 0$ is not a pole).

Straightforward computations lead to the following result for original $B(r, t)$:

$$B(r, t) = \frac{\mu_0 I}{4\pi c}\left\{\frac{J_1\left(\sqrt{\frac{-\alpha + i\omega}{\nu}}ir\right)}{J_1\left(\sqrt{\frac{-\alpha + i\omega}{\nu}}ic\right)}\exp[(-\alpha + i\omega)t]\right.$$

$$\left. + \frac{J_1\left(\sqrt{\frac{-\alpha - i\omega}{\nu}}ir\right)}{J_1\left(\sqrt{\frac{-\alpha - i\omega}{\nu}}ic\right)}\exp[-(\alpha + i\omega)t]\right. \tag{9}$$

$$+4\sum_{k=1}^{\infty}\frac{\beta_k(\beta_k^2-\zeta^2)}{(\beta_k^2-\zeta^2)^2+(\omega\tau_{\mathrm{m}})^2}\frac{J_1\left(\frac{\beta_k r}{c}\right)}{J_1'(\beta_k)}\exp\left(-\frac{\alpha\beta_k^2}{\zeta^2}t\right)\Bigg\},$$

where $\tau_{\mathrm{m}} = c^2/\nu$ is the characteristic time of magnetic field diffusion [6] and $\zeta = \sqrt{\alpha\tau_m}$.

In this expression, the first and second terms are the residues at poles s_{01} and s_{02}, respectively, and an nth term of the series is the residue at pole s_n.

Using expression (9), it is easy to derive the steady-state ($\alpha = 0$, $t \longrightarrow \infty$) magnetic field distribution in a liquid cylinder carrying a variable current of frequency ω,

$$B(r,t) = \frac{\mu_0 I}{4\pi c}\left\{\frac{J_1\left[\frac{r}{\delta}\exp\left(\frac{3\pi i}{4}\right)\right]}{J_1\left[\frac{c}{\delta}\exp\left(\frac{3\pi i}{4}\right)\right]}\exp(i\omega t) + \frac{J_1\left[\frac{r}{\delta}\exp\left(\frac{5\pi i}{4}\right)\right]}{J_1\left[\frac{c}{\delta}\exp\left(\frac{5\pi i}{4}\right)\right]}\exp(-i\omega t)\right\}, \tag{10}$$

where $\delta = \sqrt{\nu/\omega}$ is the characteristic depth of electromagnetic field penetration into the conductor [3].

Using Kelvin functions $ber_n(z)$ and $bei_n(z)$ given by [10]

$$ber_n(z) \pm i\,bei_n(z) = J_n\left[z\exp\left(\pm\frac{3\pi i}{4}\right)\right];$$

$$n = 0, 1, 2, \ldots,$$

we can represent expression (10) in the form

$$B(r,t) = \frac{\mu_0 I}{4\pi c}$$

$$\times\left\{\frac{ber_1\left(\frac{c}{\delta}\right)ber_1\left(\frac{r}{\delta}\right)+bei_1\left(\frac{c}{\delta}\right)bei_1\left(\frac{r}{\delta}\right)}{\left[ber_0'\left(\frac{c}{\delta}\right)\right]^2+\left[bei_0'\left(\frac{c}{\delta}\right)\right]^2}\cos\omega t \right. \tag{11}$$

$$\left.+\frac{bei_1\left(\frac{c}{\delta}\right)ber_1\left(\frac{r}{\delta}\right)-ber_1\left(\frac{c}{\delta}\right)bei_1\left(\frac{r}{\delta}\right)}{\left[ber_0'\left(\frac{c}{\delta}\right)\right]^2+\left[bei_0'\left(\frac{c}{\delta}\right)\right]^2}\sin\omega t\right\}.$$

Substituting magnetic field distribution (11) into Eq. (3) with regard to the recurrence relations for the Kelvin functions [11], we arrive at the following expression for the steady-state distribution of the ac current density in the liquid cylinder:

$$j(r,t) = \frac{I}{2\pi c\delta}$$

$$\times\left\{\frac{bei_0'\left(\frac{c}{\delta}\right)ber_0\left(\frac{r}{\delta}\right)-ber_0'\left(\frac{c}{\delta}\right)bei_0\left(\frac{r}{\delta}\right)}{\left[ber_0'\left(\frac{c}{\delta}\right)\right]^2+\left[bei_0'\left(\frac{c}{\delta}\right)\right]^2}\cos\omega t \right. \tag{12}$$

$$\left.-\frac{bei_0'\left(\frac{c}{\delta}\right)bei_0\left(\frac{r}{\delta}\right)+ber_0'\left(\frac{c}{\delta}\right)ber_0\left(\frac{r}{\delta}\right)}{\left[ber_0'\left(\frac{c}{\delta}\right)\right]^2+\left[bei_0'\left(\frac{c}{\delta}\right)\right]^2}\sin\omega t\right\}.$$

Up to notation, formula (12) coincides with the expression alternatively derived in [8], where the variable current distribution over the cross section of an infinitely long homogeneous cylindrical wire was calculated.

We will consider an exponentially decaying current pulse ($\omega = 0$). From expressions (3), (8), and (9), we obtain

$$\mathcal{B}(s,r) = \frac{\mu_0 I}{2\pi c(s+\alpha)}\frac{J_1\left(i\sqrt{\frac{s}{\nu}}r\right)}{J_1\left(i\sqrt{\frac{s}{\nu}}c\right)}, \tag{13}$$

$$B(r,t) = \frac{\mu_0 I}{\pi c}\left[\frac{J_1\left(\frac{\zeta r}{c}\right)}{2J_1(\zeta)}\exp(-\alpha t) + \sum_{k=1}^{\infty}\frac{\beta_k}{\beta_k^2-\zeta^2}\frac{J_1\left(\frac{\beta_k}{c}r\right)}{J_1'(\beta_k)}\exp\left(-\frac{\alpha\beta_k^2}{\zeta^2}t\right)\right], \tag{14}$$

$$j(r,t) = \frac{I}{\pi c^2}\left[\frac{\zeta}{2}\frac{J_0\left(\frac{\zeta r}{c}\right)}{J_1(\zeta)}\exp(-\alpha t) + \sum_{k=1}^{\infty}\frac{\beta_k^2}{\beta_k^2-\zeta^2}\frac{J_0\left(\frac{\beta_k r}{c}\right)}{J_1'(\beta_k)}\exp\left(-\frac{\alpha\beta_k^2}{\zeta^2}t\right)\right]. \tag{15}$$

For a current pulse decaying exponentially, quantity $\zeta^2 = \alpha\tau_m$ is the ratio of the characteristic time of magnetic field diffusion, $\tau_m = c^2/\nu$, to characteristic pulse duration $\tau_p = \alpha^{-1}$.

The distributions of the magnetic field and current density ((14) and (15), respectively) over the conductor's cross section are applicable only if all the poles of operator solution (13) are first-order poles (i.e., for $\alpha \neq s_k$, $k = 1, 2, 3, \ldots$). If, however, the equality $\zeta = \beta_n$ holds for any of indices k (e.g., for $k = n$), i.e., if point $s = -\alpha$ is a second-order pole, the inversion of expression (13) yields

$$B(r,t) = \frac{\mu_0 I}{2\pi c}\left\{\left[\frac{r}{c}\frac{J_0\left(\frac{\beta_n r}{c}\right)}{J_1'(\beta_n)} - \frac{J_1\left(\frac{\beta_n r}{c}\right)}{\beta_n J_1'(\beta_n)}\left(1+\frac{2\beta_n^2 t}{\tau_m}\right)\right]\exp\left(-\frac{\beta_n^2 t}{\tau_m}\right) + {\sum_{k=1}^{\infty}}'\frac{\beta_k}{\beta_k^2-\beta_n^2}\frac{J_1\left(\frac{\beta_k r}{c}\right)}{J_1'(\beta_k)}\exp\left(-\frac{\beta_k^2 t}{\tau_m}\right)\right\}. \tag{16}$$

Hereafter, the primed summation sign indicates that subscript $k = n$ is omitted from summation. Using expressions (16) and (3), we obtain for $\alpha = \beta_n^2/\tau_m$

$$j(r,t) = \frac{I}{2\pi c^2}\left\{\left[\frac{J_0\left(\frac{\beta_n r}{c}\right)}{J_1'(\beta_n)}\left(1-\frac{2\beta_n^2 t}{\tau_m}\right) - \frac{r}{c}\frac{\beta_n J_1\left(\frac{\beta_n r}{c}\right)}{J_1'(\beta_n)}\right]\exp\left(-\frac{\beta_n^2 t}{\tau_m}\right) + {\sum_{k=1}^{\infty}}'\frac{\beta_k^2}{\beta_k^2-\beta_n^2}\frac{J_0\left(\frac{\beta_k r}{c}\right)}{J_1'(\beta_k)}\exp\left(-\frac{\beta_k^2 t}{\tau_m}\right)\right\}. \tag{17}$$

Since zeros β_k ($k = 1, 2, 3, \ldots$) of Bessel function $J_1(u)$ increase substantially with number k [11], the series appearing in formulas (14) and (15) converge rapidly for $t > 0$ and, with time, initial terms of the series start dominating over the remaining terms. Consequently, for $\zeta \neq \beta_k$ ($k = 1, 2, 3, \ldots$), the developed stage of current pulse passage is described by simplified formulas (14) and (15), in which only the initial terms of the series that involve $\beta_1 = 3.8317$ should be taken into account, along with the terms proportional to $\exp(-\alpha t)$.

Further simplification of expressions (14) and (15) depends on the value of dimensionless parameter $\zeta = \sqrt{\tau_m/\tau_p}$. For the developed stage of the process, we have (at $\zeta < \beta_1$)

$$B(r,t) = \frac{\mu_0 I}{2\pi c}\frac{J_1\left(\frac{\zeta r}{c}\right)}{J_1(\zeta)}\exp(-\alpha t),$$
$$j(r,t) = \frac{\zeta I}{2\pi c^2}\frac{J_0\left(\frac{\zeta r}{c}\right)}{J_1(\zeta)}\exp(-\alpha t). \tag{18}$$

Taking into account these relationships and using Eq. (4) of magnetohydrostatics, we find the Lorentz force pressure distribution in the liquid conductor,

$$p_e(r,t) = p_w + \frac{\mu_0 I^2}{8\pi^2 c^2}\frac{J_0^2\left(\frac{\zeta r}{c}\right) - J_0^2(\zeta)}{J_1^2(\zeta)}\exp(-2\alpha t), \tag{19}$$

where p_w is the pressure at the tube wall ($r = c$).

In the case of a long pulse ($\varepsilon = \zeta^2 \ll 1$), the leading terms of expansions of formulas (18) and (19) in powers of small parameter ε are easy to write using the representation of the Bessel functions in the form of power series [11],

$$B(r,t) = \frac{\mu_0 r}{2}j(t), \quad j(t) = j_0\exp(-\alpha t), \quad j_0 = \frac{I}{\pi c^2},$$
$$p_e(r,t) = p_w + \frac{\mu_0 c^2 j_0^2}{4}\left[1-\left(\frac{r}{c}\right)^2\right]\exp(-2\alpha t). \tag{20}$$

The magnetic induction distribution described by the first formula in (20) is an obvious generalization of the corresponding expression for a cylindrical conductor carrying a steady-state electric current to the quasi-steady-state case.

For $\zeta = \beta_1$, expressions (16) and (17) for the developed stage of the process yield

$$\begin{aligned} B(r,t) &= \frac{\mu_0 I}{2\pi c}\left[\frac{r}{c}\frac{J_0\left(\frac{\beta_1 r}{c}\right)}{J_1'(\beta_1)} - \frac{J_1\left(\frac{\beta_1 r}{c}\right)}{\beta_1 J_1'(\beta_1)}\left(1+\frac{2\beta_1^2 t}{\tau_m}\right)\right]\exp\left(-\frac{\beta_1^2 t}{\tau_m}\right), \\ j(r,t) &= \frac{I}{2\pi c^2}\left[\frac{J_0\left(\frac{\beta_1 r}{c}\right)}{J_1'(\beta_1)}\left(1-\frac{2\beta_1^2 t}{\tau_m}\right) - \frac{r}{c}\frac{\beta_1 J_1\left(\frac{\beta_1 r}{c}\right)}{J_1'(\beta_1)}\right]\exp\left(-\frac{\beta_1^2 t}{\tau_m}\right). \end{aligned} \tag{21}$$

Substituting expressions (21) into magnetohydrostatic equation (4), we obtain

$$\begin{aligned} \frac{\partial p_e}{\partial r} &= \frac{\mu_0 I^2}{4\pi^2 c^3}\frac{\exp\left(-\frac{2\beta_1^2 t}{\tau_m}\right)}{[J_1'(\beta_1)]^2}\left[J_0\left(\frac{\beta_1 r}{c}\right)J_1\left(\frac{\beta_1 t}{c}\right)\right. \\ &\times\left(\frac{1}{\beta_1}-\frac{4\beta_1^3 t^2}{\tau_m^2}+\frac{\beta_1 r^2}{c^2}\right) - \frac{r}{c}J_0^2\left(\frac{\beta_1 r}{c}\right) \\ &\times\left.\left(1-\frac{2\beta_1^2 t}{\tau_m}\right) - \frac{r}{c}J_1^2\left(\frac{\beta_1 r}{c}\right)\left(1+\frac{2\beta_1^2 t}{\tau_m}\right)\right]. \end{aligned}$$

Integration of the second and third terms on the right of this equation using formula 5.54.2 in [12] gives

$$\begin{aligned} p_e(r,t) &= p_w + \frac{\mu_0 I^2}{4\pi^2 c^2}\frac{\exp\left(-\frac{2\beta_1^2 t}{\tau_m}\right)}{[J_1'(\beta_1)]^2} \\ &\times\left\{\frac{r}{\beta_1 c}J_0\left(\frac{\beta_1 r}{c}\right)J_1\left(\frac{\beta_1 r}{c}\right)\left(1+\frac{2\beta_1^2 t}{\tau_m}\right)\right. \\ &-\left(\frac{r}{c}\right)^2\left[J_0^2\left(\frac{\beta_1 r}{c}\right)+\frac{1}{2}J_1^2\left(\frac{\beta_1 r}{c}\right)\right] + [J_1'(\beta_1)]^2 \\ &+\left.\frac{1}{2\beta_1^2}\left(1-\frac{4\beta_1^4 t^2}{\tau_m^2}\right)\left[(J_1'(\beta_1))^2 - J_0^2\left(\frac{\beta_1 r}{c}\right)\right]\right\}. \end{aligned} \tag{22}$$

If $\zeta > \beta_1$, we proceed similarly to the case $\zeta < \beta_1$ and, using relations (4), (14), and (15), come to

$$\begin{aligned} B(r,t) &= \frac{\beta_1 \mu_0 I}{\pi c(\beta_1^2-\zeta^2)}\frac{J_1\left(\frac{\beta_1 r}{c}\right)}{J_1'(\beta_1)}\exp\left(-\frac{\alpha\beta_1^2}{\zeta^2}t\right), \\ j(r,t) &= \frac{\beta_1^2 I}{\pi c^2(\beta_1^2-\zeta^2)}\frac{J_0\left(\frac{\beta_1 r}{c}\right)}{J_1'(\beta_1)}\exp\left(-\frac{\alpha\beta_1^2}{\zeta^2}t\right), \\ p_e(r,t) &= p_w + \frac{\beta_1^2\mu_0 I^2}{2\pi^2 c^2(\beta_1^2-\zeta^2)^2} \\ &\times\left\{\frac{J_0^2\left(\frac{\beta_1 r}{c}\right)}{[J_1'(\beta_1)]^2}-1\right\}\exp\left(-\frac{2\alpha\beta_1^2}{\zeta^2}t\right). \end{aligned} \tag{23}$$

Thus, as electric current pulse $g(t) = I\exp(-\alpha t)$ propagates along the cylinder, the electromagnetic field–pressure distribution, in the general case, takes a regular, in a sense, pattern with time: the characteristic profiles of magnetic induction, current density, and pressure (see (18)–(20) and (23)) are no longer time-variable and their amplitudes decay exponentially. The exponent of the exponential is a function of dimensionless parameter $\zeta = \sqrt{\tau_m/\tau_p}$. In the special case $\zeta = \beta_1$, the regular regime does not set in, as follows from formulas (21) and (22).

CONCLUSIONS

Using the equation for the magnetic field and the Laplace transformation in time, we solved the axisymmetric problem (with a discontinuous initial condition) of electromagnetic field distribution in a liquid cylindrical conductor along which an exponentially decaying current pulse propagates.

In the special case of nondecaying variable current, the result obtained in this work coincides with the solution to the problem of skin effect [8] up to notation.

If characteristic duration τ_p of the current pulse mentioned above is much longer than characteristic time τ_m of magnetic field diffusion across the liquid cylinder, the leading term in the expansion of the solution found in small parameter $\varepsilon = \tau_m/\tau_p$ extends the formula for the magnetic field distribution in a cylindrical conductor with uniform current to the quasi-stationary case.

Using the equation of magnetohydrostatics, we found the distribution of the pressure produced by the exponentially decaying current pulse in the liquid cylinder.

In the general case, the electromagnetic field–pressure distribution becomes in a sense regular with time: the characteristic profiles of the magnetic field, current density, and pressure become time-invariable and decay exponentially. The exponents of the exponentials depend on dimensionless parameter $\zeta = \sqrt{\varepsilon}$.

In the special case $\zeta = \beta_1$, where β_1 is the first nontrivial zero of Bessel function $J_1(z)$, the regular regime does not set in.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 02-01-00694.

REFERENCES

1. S. Makarov, R. Ludwig, J. Resnik, and D. Apelian, J. Nondestruct. Eval. **18** (13), 99 (1999).
2. V. V. Boyarevich, Ya. Zh. Freiberg, E. I. Shilova, and E. V. Shcherbinin, *Electrical Eddying Flows* (Zinatne, Riga, 1985) [in Russian].
3. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1992; Pergamon, New York, 1984).
4. L. A. Artsimovich, *Controlled Thermonuclear Reactions* (Fizmatgiz, Moscow, 1961; Gordon and Breach, New York, 1964).
5. J. A. Shercliff, *A Textbook of Magnetohydrodynamics* (Pergamon, Oxford, 1965; Mir, Moscow, 1967).
6. J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975; Mir, Moscow, 1965).
7. E. I. Shilova, Magn. Gidrodinamika, No. 2, 142 (1975).
8. A. N. Tikhonov and A. A. Samarskii, *Equations of Mathematical Physics* (Nauka, Moscow, 1977; Pergamon, Oxford, 1964).
9. M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable*, 3rd ed. (Nauka, Moscow, 1973; Springer-Verlag, Berlin, 1959).
10. E. Jahnke, F. Emde, and F. Losch, *Tables of Higher Functions*, 6th ed. (McGraw-Hill, New York, 1960; Nauka, Moscow, 1964).
11. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).
12. I. S. Gradshtein and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Nauka, Moscow, 1971; Academic, New York, 1980).

Translated by N. Wadhwa

Technical Physics, Vol. 50, No. 7, 2005, pp. 821–827. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 6–12.
Original Russian Text Copyright © 2005 by Vecheslavov.

THEORETICAL
AND MATHEMATICAL PHYSICS

Contribution from the Secondary Harmonics of a Disturbance to the Separatrix Map of the Hamiltonian System

V. V. Vecheslavov

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences,
pr. Akademika Lavrent'eva 11, Novosibirsk, 630090 Russia
e-mail: vecheslavov@inp.nsk.su

Received July 6, 2004

Abstract—The special role of low-frequency secondary harmonics with frequencies that are sums of and differences between primary frequencies entering into the Hamiltonian in explicit form has been already discussed in the literature. These harmonics are of the second order of smallness and constitute a minor fraction of the disturbance. Nevertheless, under certain conditions, their contribution to the amplitude of the separatrix map of the system may be several orders of magnitude higher than the contributions from primary harmonics and, thereby, govern the formation of dynamic chaos. This work generalizes currently available theoretical and numerical data on this issue. The role of secondary harmonics is demonstrated with a pendulum the disturbance of which in the Hamiltonian is represented by two asymmetric closely spaced high-frequency harmonics. An analytical expression for the contribution of the secondary harmonics to the separatrix map amplitude for this system is derived, and the range of very low secondary frequencies not studied earlier is considered using this equation. The domains where the separatrix map amplitude linearly grows with frequency and the chaotic layer size is frequency-independent are indicated. Theoretical predictions are compared with numerical data. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

It is known that, in Hamiltonian systems with the phase space separated into regular and chaotic components, chaos forms via interaction between nonlinear resonances. One of the resonances is usually taken to be primary and specifies initial conditions, and others are considered as disturbances. The dynamics has turned out to be the most interesting and unexpected near primary resonance separatrices, special trajectories separating out domains with circulating (out of a resonance) and oscillating (within a resonance) phase. New recently discovered fine points concerning this issue deserve attention.

It was common knowledge that chaos arises just near separatrices, since the period of motion along them is infinite and the interaction of resonances has an essential effect [1–3]. It turned out, however, that such a view is adequate only if the Fourier amplitudes of the system's analytical potential decay exponentially. Here, a disturbance splits each of the separatrices into two branches, which occupy a narrow domain, forming a chaotic layer. It is remembered that the chaotic layer emerging instead of the separatrices can be subdivided into three parts: upper (phase x circulates at the top, $p > 0$), middle (the phases oscillates), and lower (phase x circulates at the bottom, $p < 0$).

In the case of a smooth potential with a power-law decay of Fourier amplitudes, the situation may become qualitatively different. Examples of unsplit separatrices (of both integer and fractional resonances) and the absence of a chaotic layer at their places in piecewise linear systems are discussed elsewhere [4–6]. It should be emphasized that the systems remain nonintegrable in this case and their separatrices are retained under strong local chaos.

It was also revealed that the formation of chaos considerably affects the disturbance itself, especially its spectral composition. Below, we will synopsize the history of this problem.

Let a system be represented by a Hamiltonian pendulum

$$H(x, p, t) = \frac{p^2}{2} + \cos(x) + V(x, t) \tag{1}$$

with a single disturbance harmonic

$$V(x, t) = \varepsilon\cos\left(\frac{n}{2}x - \Omega t\right), \tag{2}$$

where parameters n and Ω are positive. This harmonic, which is also a resonance harmonic, lies above the fundamental resonance of the system in the phase plane; therefore, it is convenient to refer to as the upper harmonic.

Chirikov [1] showed theoretically that upper disturbance harmonic (2) is responsible for amplitude W of the separatrix map of the upper part of the chaotic layer and that amplitude w obeys the relationship

$$W(\Omega, n) = \varepsilon\Omega A_n(\Omega) \tag{3}$$

where

$$A_n(\Omega > 0) = \frac{2\pi}{(n-1)!}\frac{\exp(\pi\Omega/2)}{\sinh(\pi\Omega)}(2\Omega)^{n-1}[1 + f_n(\Omega)], \qquad (4)$$

are the Mel'nikov–Arnold integrals. Functions $f_n(\Omega)$ appearing in the brackets are defined by the recurrence relations

$$f_1 = f_2 = 0,$$
$$f_{n+1} = f_n - (1 + f_{n-1})\frac{n(n-1)}{4\Omega^2}, \quad n \ge 3. \qquad (5)$$

Substitution of $-\Omega$ for Ω turns the upper harmonic into the lower one; then, in formula (3) for its contribution to the amplitude of the separatrix map of the upper part of the layer, $A_n(\Omega)$ should be taken in a much different form,

$$A_n(\Omega < 0) = (-1)^n A_n(|\Omega|)\exp(-\pi|\Omega|). \qquad (6)$$

It should be emphasized that, when deriving relationships (3)–(6), Chirikov [1] did not make any simplifying assumptions; hence, they are valid for any Ω from the interval $0 < |\Omega| < \infty$. Note also that, when Ω varies, integrals (4) pass through zeros along with the bracketed factor, the number of zeros being dependent on index n of the Mel'nikov–Arnold integral.

Chirikov's studies showed that, for a system with a symmetric disturbance of type

$$V(x, t) = \varepsilon\left[\cos\left(\frac{n}{2}x - \Omega t\right) + \cos\left(\frac{n}{2}x + \Omega t\right)\right], \qquad (7)$$

separatrix map amplitude W and the energy size of the chaotic layer,

$$w_{\text{up}} = |w_{md}| = w_{bt} = \Omega W, \qquad (8)$$

exponentially decay with increasing frequency in the high-frequency limit $\Omega \longrightarrow \infty$ and that all the three parts of the chaotic layer have the same width (in (8), $w = p^2/2 + \cos x - 1$ is the relative deviation from the undisturbed separatrix in terms of energy.

In his recent work [7], the author, considering the low-frequency ($\Omega \longrightarrow 0$) asymptotics of system (1) with symmetric disturbance (7), found that, in this limit, the separatrix map amplitude linearly grows with frequency and the width of the layer is frequency-independent. Both asymptotics are relatively simple, so that the medium-frequency range (corresponding to either a high or low adiabaticity parameter) is the most difficult to analyze. Here, the so-called resonance invariants, which fairly adequately reflect the topology of a specific resonance, proved to be of great value. For the Chirikov standard mapping, such invariants of first three orders, which correspond to 1 : 1, 1 : 2, and 1 : 3 resonances, were constructed in [8]; for the single-frequency separatrix mapping, in [9]. The expressions for double-frequency invariants, which were purposely introduced for studying chaos in the neighborhood of the zeros of the Mel'nikov–Arnold integrals, have been deduced only recently [10].

It seems that the author was the first to consider an asymmetric disturbance [11, 12]. This situation was exemplified with Hamiltonian (1) of a pendulum subjected to two harmonics with different frequencies,

$$V(x, t) = \varepsilon_1\cos(x - \Omega_1 t) + \varepsilon_2\cos(x - \Omega_2 t). \qquad (9)$$

The amplitudes of the harmonics were assumed to be small ($\varepsilon_1, \varepsilon_2 \ll 1$), while their frequencies high ($|\Omega_1|, |\Omega_2| \gg 1$).

It is in this case where secondary harmonics of order $\sim\varepsilon_1\varepsilon_2$ with frequencies

$$\Delta\Omega_+ = \Omega_1 + \Omega_2, \quad \Delta\Omega_- = \Omega_2 - \Omega_1, \qquad (10)$$

arise, which are much weaker than the primary harmonics if $\varepsilon_1, \varepsilon_2 \ll 1$ (see below).

Even early experiments with system (1) with disturbance (9) discovered the seemingly surprising fact that just these weak secondary harmonics define the separatrix map amplitude and chaotic layer size under certain conditions.

In [12], the author considered a system with parameters $\varepsilon_1 = \varepsilon_2 = 0.075$, $\Omega_1 = 13$, and $\Omega_2 = -10$, where the secondary harmonic with aggregate frequency $\Delta\Omega_+ = 3$ entering into the disturbance had an amplitude of $\approx 4.5 \times 10^{-5}$, which is ≈ 1700 times as small as the amplitudes of the primary harmonics. Yet, its contribution to the amplitude of the separatrix map of the chaotic layer upper part (this mapping is responsible for the formation of the chaotic layer) exceeds the total contribution from the primary harmonics almost 400 times. The sizes of the layer parts turned out to differ substantially. This is a consequence of the above-mentioned exponential frequency dependence of the width of the layer at $\Omega \gg 1$, which allows very weak but low-frequency secondary harmonics to have a decisive effect on formation of chaos. An important role of secondary harmonics at aggregate frequencies was also discovered in smooth systems [13].

Finally, an asymmetric disturbance in the rather general form

$$V(x, t) = \varepsilon_1\cos(m_1 x - \Omega_1 t) + \varepsilon_2\cos(m_2 x - \Omega_2 t) \qquad (11)$$

(where m_1 and m_2 are arbitrary positive integers) was considered in [14].

It is remembered that disturbance parameters m_1 and m_2 specify the structure of Mel'nikov–Arnold integrals (4) and (5), in particular, the number of zeros. In the neighborhood of these zeros, dynamic chaos has a number of specific features, which were discussed in [14, Sect. 3]. Below, we will consider system (1) with disturbance (11) again. Note in passing that we will deal with integrals (4) and (5) that have only even indices n, since m_1 and m_2 are integers.

SECONDARY-HARMONIC AMPLITUDES

The amplitudes of disturbance secondary harmonics are not known in advance and are to be determined. A relevant rigorous theory is still lacking; however, a general approach to tackling the problem and finding approximate analytical relationships was suggested in [11]. Following [11], we change variables in (1) and (11); namely, instead of coordinate $x(t)$ and momentum $p(t)$, we will use deviations from their values $x_s(t) = 4\arctan(e^t)$ and $p_s(t) = 2\sin(x_s(t))$ on the undisturbed separatrix,

$$\begin{aligned} y(t) &= x(t) - x_s(t), \\ u(t) &= p(t) - p_s(t). \end{aligned} \tag{12}$$

Then, using generating function $F_2(u, x, t) = [p_s(t) - u][x - x_s(t)]$, we will construct new Hamiltonian

$$\begin{aligned} H(y, u, t) &= \frac{u^2}{2} + \cos y \cos x_s(t) - \sin y \sin x_s(t) \\ &+ y \sin x_s(t) + \sum_{k=1}^{2} \varepsilon_k [\cos(m_k y)\cos(m_k x_s(t) - \Omega_k t) \\ &- \sin(m_k y)\sin(m_k x_s(t) - \Omega_k t)]. \end{aligned} \tag{13}$$

Since the disturbance is small, we put $|y(t)| \ll 1$ and change variables in (13), $\cos(my) \longrightarrow 1 - (my)^2/2$ and $\sin(my) \longrightarrow my$, to arrive at the equation

$$\begin{aligned} d^2y/dt^2 &= y\left[\cos x_s + \sum_{k=1}^{2} \varepsilon_k m_k^2 \cos(m_k x_s - \Omega_k t)\right] \\ &+ \sum_{k=1}^{2} \varepsilon_k m_k \sin(m_k x_s - \Omega_k t). \end{aligned}$$

Let Δy_ε be the difference between the left- and right-hand sides of this equation,

$$\begin{aligned} \Delta y_\varepsilon &= \frac{d^2y}{dt^2} - y\left[\cos x_s + \sum_{k=1}^{2} \varepsilon_k m_k^2 \cos(m_k x_s - \Omega_k t)\right] \\ &- \sum_{k=1}^{2} \varepsilon_k m_k \sin(m_k x_s - \Omega_k t). \end{aligned} \tag{14}$$

We are interested in only "induced" solution y_ε (i.e., that vanishing at $\varepsilon \longrightarrow 0$), which can be conveniently found by the method of successive approximations in the limit $\Delta y_\varepsilon \longrightarrow 0$ [11]. After two approximations, we have

$$\begin{aligned} y_\varepsilon^{(2)}(t) &\approx -\sum_{k=1}^{2} \frac{\varepsilon_k m_k}{(m_k p_s - \Omega_k)^2} \sin(m_k x_s - \Omega_k t) \\ &- \frac{\varepsilon_1 \varepsilon_2 m_1 m_2}{2} \left\{ \left[\frac{m_2}{(m_1 p_s - \Omega_1)^2} + \frac{m_1}{(m_2 p_s - \Omega_2)^2}\right] \right. \\ &\times \frac{\sin(m_+ x_s - \Delta\Omega_+ t)}{(m_+ p_s - \Delta\Omega_+)^2} \\ &+ \left[\frac{m_1}{(m_2 p_s - \Omega_2)^2} - \frac{m_2}{(m_1 p_s - \Omega_1)^2}\right] \\ &\left. \times \frac{\sin(m_- x_s - \Delta\Omega_- t)}{(m_- p_s - \Delta\Omega_-)^2} \right\} + \ldots, \end{aligned} \tag{15}$$

where minor terms are omitted and $m_+ = m_1 + m_2$ and $m_- = m_2 - m_1$.

In [14], attention was focused mainly on the neighborhoods of the zeros of integrals (4), which are expandable when $\Delta\Omega_\pm$ is on the order of unity; therefore, in [14], to simplify the resulting expressions, the terms proportional to mp_s were omitted in all the denominators of formula (16) (here, formula (15)). In this work, we will concentrate on the case of very low secondary frequencies, $\Delta\Omega_\pm \ll 1$ (see next section); therefore, we dismiss these simplifications and use rigorous formulas.

Turning back to system (1) with disturbance (11), we put in it $x = x_s(t) + y_\varepsilon^{(2)}(t)$. Since the motion is near the undisturbed separatrix, we make change of variables, $\cos my \longrightarrow 1$ and $\sin my \approx my_\varepsilon^{(2)}$, and recast expression (11) for the disturbance as

$$V(y, t) \approx -y_\varepsilon^{(2)}\left[\sin x_s + \sum_{k=1}^{2} \varepsilon_k m_k \sin(m_k x_s - \Omega_k t)\right]. \tag{16}$$

Substituting (15) into (16) shows that the low-frequency harmonics at the aggregate and difference frequencies (which alone are of interest) may arise in the disturbance in two ways. One is the interaction of sum (15) with primary harmonics in (16), which gives rise to harmonics at both the aggregate frequency,

$$\begin{aligned} &\varepsilon_+ \cos(m_+ x_s - \Delta\Omega_+ t), \\ \varepsilon_+ &= -\frac{\varepsilon_1 \varepsilon_2 m_1 m_2}{2}\left[\frac{m_2}{(m_1 p_s - \Omega_1)^2} + \frac{m_1}{(m_2 p_s - \Omega_2)^2}\right], \end{aligned} \tag{17}$$

and difference frequency,

$$\varepsilon_{-}[\cos(m_{-}x_{\mathrm{s}} - \Delta\Omega_{-}t)],$$

$$\varepsilon_{-} = -\frac{\varepsilon_1\varepsilon_2 m_1 m_2}{2}\left[\frac{m_1}{(m_2 p_{\mathrm{s}} - \Omega_2)^2} - \frac{m_2}{(m_1 p_{\mathrm{s}} - \Omega_1)^2}\right]. \quad (18)$$

Note that formulas (17) and (18) describe the contributions from harmonic–harmonic interaction.

The other way is the interaction of terms $\sim\varepsilon_1\varepsilon_2$ in (15) with the term $\sin x_{\mathrm{s}}$ in (16), which is born by the fundamental resonance. Here, two harmonics arise at the aggregate frequency,

$$\frac{\varepsilon_{+}}{2}\left\{\frac{\cos((m_{+}-1)x_{\mathrm{s}} - \Delta\Omega_{+}t)}{[(m_{+}-1)p_{\mathrm{s}} - \Delta\Omega_{+}]^2} - \frac{\cos((m_{+}+1)x_{\mathrm{s}} - \Delta\Omega_{+}t)}{[(m_{+}+1)p_{\mathrm{s}} - \Delta\Omega_{+}]^2}\right\} \quad (19)$$

(ε_{+} is given by (17)), and two at the difference frequency,

$$\frac{\varepsilon_{-}}{2}\left\{\frac{\cos((m_{-}-1)x_{\mathrm{s}} - \Delta\Omega_{-}t)}{[(m_{-}-1)p_{\mathrm{s}} - \Delta\Omega_{-}]^2} - \frac{\cos((m_{-}+1)x_{\mathrm{s}} - \Delta\Omega_{-}t)}{[(m_{-}+1)p_{\mathrm{s}} - \Delta\Omega_{-}]^2}\right\} \quad (20)$$

(where ε_{-} is given by (18)).

Formulas (19) and (20) describe the contributions due to interaction between the disturbance-induced terms and fundamental resonance.

Knowing the amplitudes of secondary harmonics entering into the disturbance, one can, with (3), represent the amplitudes of their separatrix maps through Mel'nikov–Arnold integrals (4),

$$W_{\pm} = -\tilde{\varepsilon}_{\pm}a_{\pm}\Delta\Omega_{\pm}\left[A_{2m_{\pm}}(\Delta\Omega_{\pm}) + \frac{A_{2m_{\pm}-2}(\Delta\Omega_{\pm})}{2[(m_{\pm}-1)p_{\mathrm{s}} - \Delta\Omega_{\pm}]^2} - \frac{A_{2m_{\pm}+2}(\Delta\Omega_{\pm})}{2[(m_{\pm}+1)p_{\mathrm{s}} - \Delta\Omega_{\pm}]^2}\right], \quad (21)$$

$$\tilde{\varepsilon}_{\pm} = \frac{m_1 m_2 \varepsilon_1 \varepsilon_2}{2}\left[\frac{m_1}{[m_2 p_{\mathrm{s}} - \Omega_2]^2} \pm \frac{m_2}{[m_1 p_{\mathrm{s}} - \Omega_1]^2}\right]. \quad (22)$$

Here, the plus and minus signs in the subscripts refer to the aggregate and difference frequencies, respectively, and a_{+} and a_{-} in relationships (21) are empiric adjusting coefficients.

Experience in applying formula (21) suggests that the first term in the brackets play a dominant role.

As was mentioned above, no simplifications were made in deriving formulas (21) and (22); hence they are valid for any primary-harmonic frequencies [14]. In the next section, we will consider at length the case of a very-low-frequency secondary harmonic (arising at the sum of primary-harmonic frequencies that are close in absolute value), where simplifications seem to be appropriate.

THE CASE OF CLOSELY SPACED PRIMARY HARMONICS

Suppose that a disturbance in the Hamiltonian involves two frequencies close in absolute value and opposite in sign, $\Omega_1 \approx |\Omega_2| \gg 1$. Here, of interest is only the low-frequency secondary harmonic at aggregate frequency $\Delta\Omega_{+} = \Omega_1 + \Omega_2 \ll 1$. The secondary harmonic at the difference frequency is much weaker (or does not arise at all if the disturbance parameters are equal to each other, $m_1 = m_2$; see (22)). Therefore, the sign minus in the subscripts will be omitted.

Consider now Mel'nikov–Arnold integrals (4) in the low-frequency limit, $\Delta\Omega_{+} \longrightarrow 0$. Here, the asymptotic expression for the bracketed sum is essential. For even $n = 2, 4, 6, \ldots$, it can be represented as

$$1 + f_n(\Delta\Omega_{+}) \longrightarrow s_n(\Delta\Omega_{+}) = \frac{c_n}{\Delta\Omega_{+}^{n-2}}, \quad \Delta\Omega_{+} \longrightarrow 0, \quad (23)$$

where coefficients c_n are found by recurrence relations (5),

$$c_2 = 1, \quad c_4 = -2, \quad c_6 = \frac{23}{2}, \quad c_8 = -132, \quad c_{10} = \frac{5067}{2}\ldots. \quad (24)$$

With (24), it is easy to check that integrals (4) become frequency-independent in the low-frequency limit and tend to some constants K_n that depend only on n,

$$K_n = A_n(\Omega \longrightarrow 0) = \frac{2^n}{(n-1)!}c_n. \quad (25)$$

Thus, the separatrix map amplitude in the low-frequency limit becomes a linear function of frequency,

$$W_{+} = -\tilde{\varepsilon}_{+}\Delta\Omega_{+}\left[K_{2m_{+}} + \frac{K_{2m_{+}-2}}{2[(m_{+}-1)p_{\mathrm{s}}]^2} - \frac{K_{2m_{+}+2}}{2[(m_{+}+1)p_{\mathrm{s}}]^2}\right] \approx -\tilde{\varepsilon}_{+}C_{+}\Delta\Omega_{+}. \quad (26)$$

It was shown [7] that, when the separatrix map amplitude varies by a linear law, the size of the upper part of the chaotic layer becomes frequency-independent. This statement will be verified in the next section; here, it seems reasonable to recall the essence of separatrix mapping.

This mapping, first introduced in [15], describes the behavior of a Hamiltonian system moving near the separatrix of the fundamental resonance. In the case of a pendulum, separatrix mapping is stated as

$$\bar{w} = w + \sum_l W_l \sin\Omega_l t_\pi,$$
$$\bar{t_\pi} = t_\pi + \ln\left(\frac{32}{|\bar{w}|}\right); \quad l = 1, 2, \ldots, \tag{27}$$

where $w = p^2/2 + \cos x - 1$ is the relative energy deviation from the undisturbed separatrix (see (8)) and t_π are time instants the system passes through the positions of stable equilibrium ($x = \pi$).

The summation sign must cover all harmonics that are significant for a chaotic layer part under study, both primary (explicitly entering into disturbance (11)) and secondary (which are absent in (11)).

If frequencies Ω_l are incommensurate, time instants t_π are counted on the continuous time scale. If the frequencies are multiples of some reference frequency Ω_0, the second relationship in (27) may be recast as

$$\bar{\psi_\pi} = \psi_\pi + \Omega_0 \ln\left(\frac{32}{|\bar{w}|}\right),$$
$$\psi_\pi = \Omega_0 t_\pi \quad (\mathrm{mod}\, 2\pi). \tag{28}$$

The quickest way of determining the sizes of the parts of the chaotic layer is iteration of the separatrix map; therefore, efforts spent on separatrix mapping prove out.

Our goal is to compare the separatrix map amplitudes found analytically and numerically. Prior to doing this, an algorithm for numerically constructing the map is worth outlining (for details, see [11]). First, a central homoclinic point p_{hb} is found on the line of symmetry $x = \pi$ with a high accuracy, this point being the boundary between phase circulation and oscillation. In the vicinity of this point, a narrow momentum interval $x = \pi$, $p_{\mathrm{hb}} < p < p_{\mathrm{hb}} + \delta p$ is taken from which a random trajectory is issued. This trajectory either executes a preset number of periods of motion or is interrupted passing to other part of the layer. In both cases, another random trajectory is issued from the same interval and so on until the number of periods reaches a desired value N. For each of the periods, an energy deviation from the undisturbed separatrix is determined,

$$w = 32\exp(-T), \tag{29}$$

where T is the time interval between two successive passings through the position of stable equilibrium $x = \pi$.

Determining a change in the energy deviation, $\delta w = \bar{w} - w$, for each pair of successive periods and assigning it to time instant t_π shared by this pair of periods, one can construct separatrix map $(\delta w)_k$, $t_{\pi,k}$, where $k = 1, 2, \ldots, N - 1$ (see (27)). To be definite, we will study the upper part of the chaotic layer. The outer (upper and lower) parts of the layer are of special interest, since they are responsible for resonance overlap and global chaos formation.

THEORETICAL RESULTS VERSUS NUMERICAL DATA

As was noted above, the neighborhoods of the zeros of Mel'nikov–Arnold integrals were considered in [14] and so here they are not discussed. Therefore, in this section, we will consider disturbance (11) with fixed parameters,

$$\varepsilon_1 = \varepsilon_2 = 0.01, \quad m_1 = 1, \quad m_2 = 1 \tag{30}$$

and different primary-harmonic frequencies Ω_1 and Ω_2. Note that, with the disturbance thus selected, Mel'nikov–Arnold integrals (4) do not have zeros throughout the frequency range $0 < \Omega < \infty$.

To handle quantities on the order of unity, we will use the reduced separatrix map amplitude and reduced energy size of the upper part of the chaotic layer,

$$W^* = \frac{10^3 W}{\varepsilon_1\varepsilon_2}, \quad w^*_{\mathrm{up}} = \frac{10^3 w_{\mathrm{up}}}{\varepsilon_1\varepsilon_2}. \tag{31}$$

To begin with, consider the (symmetric) case of frequencies that are equal in magnitude and opposite in sign,

$$\Omega_1 = 16.0, \quad \Omega_2 = -16.0. \tag{32}$$

The separatrix map amplitude and the half-width of the layer calculated by Chirikov's formulas (3)–(6) were found to be $W^* = 3.92 \times 10^{-2}$ and $w^*_{\mathrm{up}} \approx 0.65$, respectively. Numerical construction of the map by formulas (27) and (28) and subsequent iterations confirm

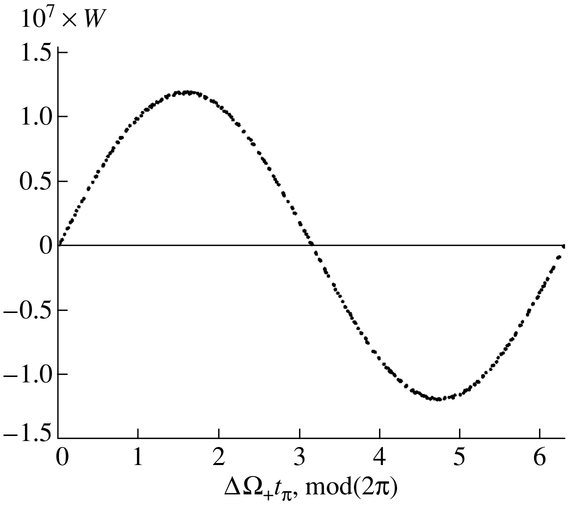


Fig. 1. Numerical construction of separatrix maps (27) and (28) for system (1) with disturbance (11) for frequencies $\Omega_1 = 16.0$, $\Omega_2 = -15.9$, and $\Delta\Omega_+ = 0.1$. $W_{\max} = 1.20 \times 10^{-7}$.

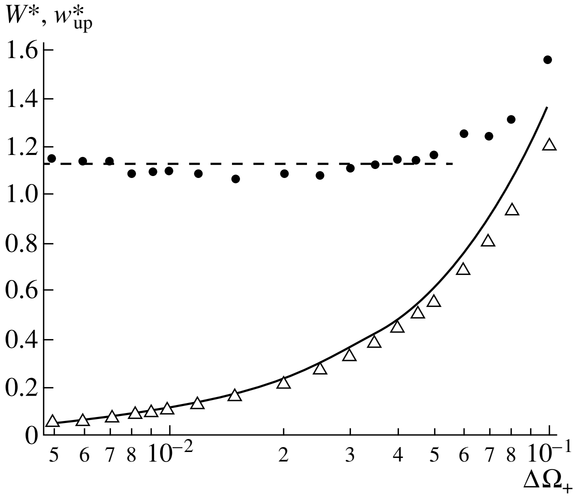


Fig. 2. System (1) with disturbance (11) for $\Omega_1 = 16.0$, Ω_2 = var, and $\Delta\Omega_+ > 0$. The continuous curve, the run of the reduced separatrix map amplitude W^* calculated by formula (21) at $a_+ = 0.4$; (△) numerical calculation of W^*; (●) numerical iteration of separatrix map (27) of reduced size $w^*_{\rm up}$ of the chaotic layer upper part; and the dashed line, domain where $w^*_{\rm up} \approx$ const.

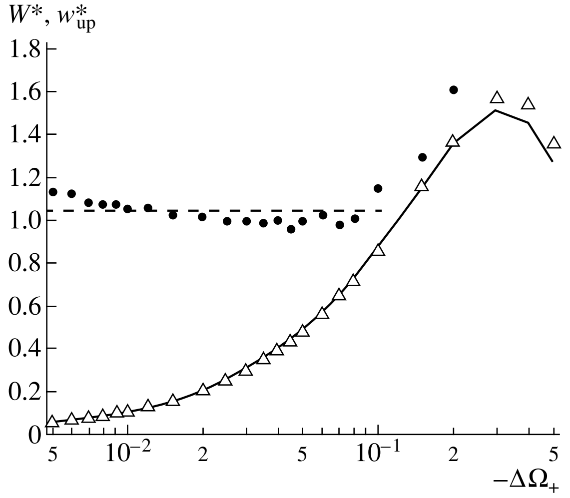


Fig. 3. System (1) with disturbance (11) for $\Omega_1 = 16.0$, Ω_2 = var, and $\Delta\Omega_+ < 0$. The designations are the same as in Fig. 2.

these values, so that the pattern has the form of a sinusoid with frequency $\Omega = 16.0$.

Seemingly, that a small change in one of the frequencies cannot dramatically change the result obtained. This assumption can be checked by making the disturbance slightly asymmetric,

$$\Omega_1 = 16.0, \quad \Omega_2 = -15.9, \quad \Delta\Omega_+ = 0.1. \qquad (33)$$

Figure 1 shows the results of numerical simulation. The separatrix map is seen to remain single-frequency, but its amplitude increased by roughly 30 times (W^* = 1.20) and, more importantly, the frequency of the map, initially high ($\Omega = 16.0$) became very low ($\Delta\Omega_+ = 0.1$). The upper part of the layer expanded by more than twice, $w^*_{\rm up} \approx 1.55$. Note in passing that the total contribution of the primary harmonics of the disturbance to amplitude W is as low as less than 2%. Here, we again face the situation where the formation of chaos in the upper part of the layer fully depends on the weak but low-frequency secondary harmonic [11, 12].

Studying the low-frequency limit, $\Omega \longrightarrow 0$, for a pendulum subjected to a symmetric disturbance [7], the author discovered domains where the separatrix map amplitude is a linear function of frequency and the size of the upper part of the chaotic layer is frequency-independent. However, in [7], low frequencies were assigned to primary harmonics, which entered into the Hamiltonian in explicit form (such a case is rarely encountered in practice), and secondary ones did not arise at all. The situation considered in this work is qualitatively different: the disturbance involves both high-frequency primary and low-frequency secondary harmonics. It is of interest to see how the above dependences change when the sum of high primary frequencies tends to zero.

Figure 2 and 3 demonstrate the behavior of the reduced (see (31)) separatrix map amplitude and reduced size of the upper size of the chaotic layer for the upper and lower harmonics (the superscript "+" in W^*_+ is omitted). Good agreement between the theory and calculation is obvious. We see that, when the disturbance is symmetric, the limit $\Delta\Omega_+ \longrightarrow 0$ for secondary harmonics leads to the same dependences as for low-frequency primary ones: the separatrix map amplitude linearly grows with frequency, while the size of the layer varies with frequency insignificantly. At the same time, it is noteworthy that the frequency independence of the size of the chaotic layer predicted by the theory is approximate and is markedly violated at the extremes of the interval. At the right extreme, the growth of the layer is naturally related to the fact that the curve leaves the range of low-frequency asymptotics. At the left, the contribution from the secondary harmonic to the separatrix map amplitude becomes comparable to that from the primary harmonic and the map is no longer single-frequency.

CONCLUSIONS

The above-mentioned studies of dynamic chaos formation in Hamiltonian systems widely exploit the model of a pendulum subjected to different harmonic disturbances. This model, along with its discrete analogue, the Chirikov standard map, is very popular in investigations into nonlinear dynamics.

The case when the disturbance in the Hamiltonian is represented by two harmonics with different frequencies is considered most comprehensively. The theoretical results and experimental data reported suggest that secondary harmonics with frequencies that are the sum of or difference between primary-harmonic frequencies are real objects. Under certain conditions, it is the secondary harmonics that govern the formation of the chaotic component of motion.

However, the analytical relationships presented in this work are approximate and need empirical adjusting coefficients to fit numerical experiment data (see formula (21)); the essence of these coefficients yet to be understood. The author believes that elaboration of an exhaustive theory of Hamiltonian systems needs additional experimental and theoretical insight into the role of secondary harmonics in formation of dynamic chaos.

ACKNOWLEDGMENTS

This work was partially supported by the research program "Mathematical Methods in Nonlinear Dynamics" at the Russian Academy of Sciences.

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
2. A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer-Verlag, New York, 1982; Mir, Moscow, 1984).
3. G. M. Zaslavskii and R. Z. Sagdeev, *Introduction to Nonlinear Physics: From Pendulum to Turbulence and Chaos* (Nauka, Moscow, 1988; Harwood, Chur, 1988).
4. S. Bullett, Commun. Math. Phys. **107**, 241 (1986).
5. V. V. Vecheslavov, Preprint No. 2000-27, IYaF (Institute of Nuclear Physics, SD RAS, Novosibirsk, 2000); nlin.CD/0005048.
6. V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **120**, 740 (2001) [JETP **93**, 649 (2001)].
7. V. V. Vecheslavov, Zh. Tekh. Fiz. **74** (5), 1 (2004) [Tech. Phys. **49**, 521 (2004)].
8. V. V. Vecheslavov, Zh. Tekh. Fiz. **58** (1), 20 (1988) [Sov. Phys. Tech. Phys. **33**, 11 (1988)].
9. V. V. Vecheslavov, Zh. Tekh. Fiz. **72** (2), 20 (2002) [Tech. Phys. **47**, 160 (2002)].
10. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **125**, 399 (2004) [JETP **98**, 352 (2004)].
11. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].
12. V. V. Vecheslavov, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 989 (1996) [JETP Lett. **63**, 1047 (1996)].
13. V. V. Vecheslavov, Zh. Tekh. Fiz. **73** (9), 1 (2003) [Tech. Phys. **48**, 1079 (2003)].
14. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **126**, 537 (2004) [JETP **99**, 663 (2004)].
15. G. M. Zaslavskii and N. N. Filonenko, Zh. Éksp. Teor. Fiz. **54**, 1590 (1965) [Sov. Phys. JETP **27**, 851 (1965)].

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 828–834. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 13–18.
Original Russian Text Copyright © 2005 by Sidnyaev.

GASES
AND LIQUIDS

Nonstationary Heat Conduction Over a Blunt Body with Surface Mass Exchange Placed in a Supersonic Flow

N. I. Sidnyaev

Bauman State Technical University,
Vtoraya Baumanskaya ul. 5, Moscow, 107005 Russia
e-mail: Sidn_ni@mail.ru

Received October 29, 2004

Abstract—Interaction of a high-temperature gas with a heat-reflecting coating is accompanied by many interrelated processes. The need for thermal protection arises when an unprotected structure is bound to inevitably fail under the action of heat fluxes. It seems that heat fluxes on the order of 2.5×10^5 W/m^2, which corresponds to equilibrium surface temperatures exceeding 1500 K, set the upper operating temperature limit for unprotected refractory metals. However, this limit is to some extent conditional, since mechanical and corrosion effects may often aggravate the thermal action, causing the structure to fail at much lower temperatures. *© 2005 Pleiades Publishing, Inc.*

(1) Generally, this problem should be tackled by solving a set of differential equations describing transient heat and mass transfer in a gas–solid system [2, 4]. Elucidation of a failure mechanism, i.e., insight into elementary physicochemical processes responsible for the failure of heat-reflecting materials, is a great challenge. Only a few of the materials that have a high heat of sublimation can actually be used as heat-reflecting materials. For example, the low heat-absorption efficiencies of metals and ice are due to the fact that, when melting, they form a low-viscosity film of the melt, which is instantly blown off the surface by the incoming gas flow [3]. Strictly speaking, destructible heat-reflecting systems perform two functions: they absorb heat and simultaneously block the incident heat flux by injecting the gas into the boundary layer (as is the case of cooling by mass exchange). It is important to emphasize that the idea of destructible heat protection implies some phase or (in the general case) physicochemical transformation taking part of the material to the gaseous state. Note also that analysis of transient heat transfer in the heat-reflecting coating is necessary in two cases: (i) for determining the thickness of the coating and (ii) for calculating the temperature of the outer surface (and, at the same time, the thickness of the layer removed if the material fails on this surface).

(2) Let us consider the flow of a high-temperature gas about a hemispherical body of surface area S and wall thickness L. Under the action of the steady high-temperature flow, the coating fails and sublimates from the surface. Each kilogram of the sublimating material carries away a certain amount of heat ΔQ. The thermal and physical properties of the material are assumed to be invariable. Physicochemical transformations in the bulk of the material and emission from the outer surface are absent. In the general case, the convective and radiant heat fluxes arriving at the surface of the coating from the outside are absorbed in the following scheme [1]:

$$\begin{aligned} q_0 + q_R &= \frac{\alpha}{c_p}(I_e - I_\mathrm{w}) + q_R \\ &= G_\Sigma \Delta Q_\mathrm{mel} + G_w \Delta Q_w + \varepsilon\sigma T_\mathrm{w}^4 + q_\lambda + q_\mathrm{inj}, \end{aligned} \tag{1}$$

where $G_\Sigma \Delta Q_\mathrm{mel}$ is the thermal effect of melting, $G_\mathrm{w}\Delta Q_\mathrm{w}$ is the thermal effect of physicochemical transformations at the wall surface, $\varepsilon\sigma T_\mathrm{w}^4$ is the radiation from the walls, q_λ is the thermal flux heating the inner layers, and q_inj is the heat removed by gas injection.

According to the Fourier law,

$$q_\lambda = -\lambda \left.\frac{\partial T}{\partial n}\right|_\mathrm{w}, \tag{2}$$

where n is the normal to the surface of the body.

The net mass flow in Eq. (1), $G_\Sigma = G_\mathrm{w} + G_\mathrm{g} + G_\mathrm{mel}$, is the sum of the mass flow of the material formed during surface transformations, the mass flow of the gas liberated from the coating, and the mass flow of the material being carried away with the film of the melt. For a certain known law of variation of heat flux, $q_0(\tau)$, performance analysis of a destructible heat-protecting coating consists of three stages: (i) determination of the heating time to failure, (ii) calculation of the thickness of the layer blown off, and (iii) determination of the depth of the heated zone after the heat flux and, hence, mass removal from the outer surface have been terminated. Under the action of the heat flux on the heat-reflecting coating, direct solid–gas transformation of the material may take place. If this process occurs on the surface exposed to the heat flux, we are dealing with a sublimating coating.

(3) Let us consider the 2D problem of heat transfer in a homogeneous cylindrical sheath of radius R and thickness L (Fig. 1). The outer surface steadily erodes at a constant temperature under the action of a heat flux directed normally to the surface, and the inner surface is just maintained at a constant temperature. If the thickness of the sheath is much smaller than its inner radius (i.e., $L \ll R$), heating is uniform; i.e., the 2D problem reduces to the 1D case. The depth of heating (the distance normal to the wall along which the temperature difference is $T_w - T_0$) is assumed to be substantially smaller than the distance along which the same temperature difference takes place on the surface. Note that this assumption is invalid if heat flux q changes sharply over the surface or the thermal conductivity of the material is high.

To solve the problem posed, it is necessary to apply a more general equation of conservation of energy in the condensed phase (heat conduction equation) in which processes due to heat absorption and/or evolution inside the heat-reflecting layer are taken into account.

A physical equivalent of the given computational model is melting or sublimation of crystals under intense aerodynamic heating [3, 5]. The heating process itself can be divided into two stages. At the first stage ($\tau < \tau_T$), the surface temperature increases monotonically until it reaches failure temperature T_f. At this temperature, crystals intensely evaporate and the vapor is immediately taken away by the incoming gas. We assume that the failure temperature and its associated thermal effect ΔQ remain constant.

The value of T_f depends on the failure mechanism in a given heat-reflecting material [1, 3]. Obviously, this temperature may not be attained under certain heating conditions. For invariable thermal and physical properties of the material and in the absence of internal physicochemical transformations, the temperature field is described by the classical heat conduction equation [6, 7]

$$\frac{\partial T(t,z)}{\partial t} = a\frac{\partial^2 T(t,z)}{\partial z^2}, \tag{3}$$

where $a = \lambda/c\rho$ is the thermal diffusivity.

Disregarding radiation from the outer surface, injection, and fusion, we can reduce heat balance equation (1) to the simplest boundary condition

$$q_0(t) = q_\lambda = -\lambda\frac{\partial T}{\partial z}\bigg|_{z=0}. \tag{4}$$

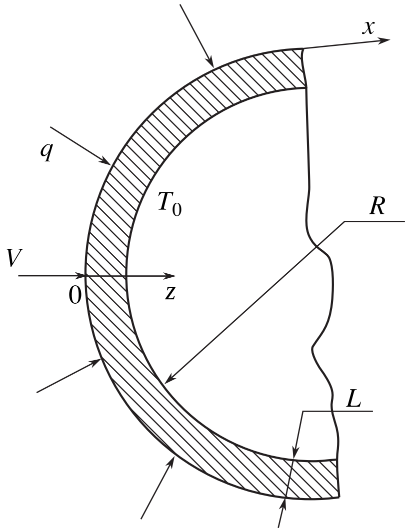


Fig. 1. Spherical sheath exposed to normally incident heat flux.

For the second boundary condition, we take

$$T(z=0) = T_0.$$

This condition meets the requirements for heat-reflecting coatings used in practice, since the sheath they coat must be kept low temperatures most of the time under transient heating [3]. As the initial condition, we take the equality of the temperature inside the coating to T_0.

Heating depth δ_T in a heat-reflecting coating is defined as the distance from the failure surface to a certain isothermal surface [6] whose temperature T_1 meets the condition $T_1 - T_0 = 0.1(T_f - T_0)$.

(4) If the failure temperature is reached, the solution of the problem reduces to solving the 1D nonstationary heat conduction equation

$$\frac{\partial T(t,z)}{\partial t} = a\frac{\partial^2 T(t,z)}{\partial z^2} \tag{5}$$

(where $a = \lambda/c\rho$) with the boundary conditions

$$\begin{aligned} -\lambda\frac{\partial T(t,z)}{\partial z} &= q \quad \text{for} \quad z = 0,\\ T(t,z) &= T_C \quad \text{for} \quad z = L, \end{aligned} \tag{6}$$

and the initial condition

$$T(0,z) = T_0, \quad T_0 = T_C. \tag{7}$$

The inhomogeneity of the boundary conditions does not allow us to seek a solution to the problem with the Fourier method. Consequently, we must use the integral

Laplace transformation in variable t [7],

$$\tilde{T}(p, z) = \int_0^{\infty} T(t, z)\exp(-pt)dt, \tag{8}$$

where Laplace transform $\tilde{T}(p, z)$ depends on z and parameter p (the latter is complex in our case).

It should be noted that the problem splits into two subproblems: finding a solution in transforms and then returning to the original.

Applying transformation (8) with the initial condition to Eq. (5) with conditions (6) and (7), we obtain the expression

$$p\tilde{T}(p, z) - T_0 = a\frac{\partial^2\tilde{T}(p, z)}{\partial z^2}$$

for transform $\tilde{T}(p, z)$ with the boundary conditions

$$-\lambda\frac{\partial\tilde{T}(p, z)}{\partial z} = \frac{q_\lambda}{p} \quad \text{for} \quad z = 0,$$

$$\tilde{T}(p, z) = \frac{T_C}{p} \quad \text{for} \quad z = L$$

and the initial conditions

$$T(0, z) = 0, \quad T_0 = T_C.$$

The solution to this problem has the form

$$\tilde{T}(p, z) = \frac{T_0}{p} + C_1\sinh\left(\sqrt{\frac{p}{a}}z\right) + C_2\cosh\left(\sqrt{\frac{p}{a}}z\right).$$

Substituting the boundary conditions yields the set of linear algebraic equations

$$\begin{cases} \sqrt{\dfrac{p}{a}}(C_1) = -\dfrac{\lambda q_\lambda}{p} \\ C_1\sinh\left(\sqrt{\dfrac{p}{a}}L\right) + C_2\cosh\left(\sqrt{\dfrac{p}{a}}L\right) = 0, \end{cases}$$

$$\begin{cases} C_1 = -\dfrac{\lambda q_\lambda}{p}\sqrt{\dfrac{a}{p}} \\ C_2 = \dfrac{\lambda q_\lambda}{p}\sqrt{\dfrac{a}{p}}\tanh\left(\sqrt{\dfrac{p}{a}}L\right). \end{cases}$$

Thus, the desired solution in transforms has the form

$$\tilde{T}(p, z) = \frac{T_0}{p} + \frac{q_\lambda}{\lambda p}\sqrt{\frac{a}{p}}\left(\frac{\sinh\left(\sqrt{\dfrac{p}{a}}(L - z)\right)}{\cosh\left(\sqrt{\dfrac{p}{a}}L\right)}\right).$$

Now we return to the original of the solution, using the standard theorems of operational calculus [7].

In accordance with the third expansion theorem, the original of the desired solution is equal to the sum of the residues of function $\tilde{T}(p, z)$ in its poles. The poles of function $\tilde{T}(p, z)$ are the roots of the equation

$$p\cosh\left(\sqrt{\frac{p}{a}}L\right) = 0.$$

These roots are $p_0 = 0, p_1, \ldots, p_k$ ($k \in N$). In this case,

$$T(t, z) = \sum_{n=0}^{\infty}\operatorname*{Res}_{p=p_n}(\tilde{T}(p, z)\exp(pt)).$$

The root $p_0 = 0$ deserves special consideration. Using the definition of residue to calculate a residue at this point, we can write

$$\operatorname*{Res}_{p=p_n}(f(p)) = \lim_{p \to p_n} f(p)(p - p_n).$$

As applied to the given case, this equality assumes the form

$$\operatorname*{Res}_{p=0}(\tilde{T}(p, z)\exp(pt)) = \lim_{p \to 0}\{\exp(pt)p\tilde{T}(p, z)\}$$

$$= T_0 + \frac{q_\lambda}{\lambda}\lim_{p \to 0}\left\{\frac{1}{\cosh\left(\sqrt{\dfrac{p}{a}}L\right)}\left(\frac{\sinh\left(\sqrt{\dfrac{p}{a}}(L - z)\right)}{\sqrt{\dfrac{p}{a}}}\right)\right\}$$

$$= T_0 + \frac{q_\lambda}{\lambda}(L - z).$$

The equation

$$\cosh\left(\sqrt{\frac{p}{a}}L\right) = 0$$

has no real roots. Therefore, making the substitution $\sqrt{p} = i\mu$, we get

$$\cos\left(\frac{L\mu}{\sqrt{a}}\right) = 0.$$

This equation has a countable set of simple roots

$$\mu_k = \frac{\sqrt{a}}{L}\left(\frac{\pi}{2} + \pi k\right), \quad k \in N.$$

They all are simple poles, so that the residue in them can be calculated by the formula

$$\operatorname*{Res}_{p=p_n}(f(p)) = \frac{\varphi(p_n)}{\psi'(p_n)},$$

where

$$\varphi(p_n) \neq 0, \quad \psi(p_n) = 0, \quad \psi'(p_n) \neq 0,$$

$$\operatorname*{Res}_{p = p_n} s\tilde{T}(p, z)\exp(pt) = \frac{\dfrac{2qa}{L\mu_n^2}\sinh\left(\dfrac{i\mu_n}{\sqrt{a}}(L-z)\right)}{\sinh\left(\dfrac{iL\mu_n}{\sqrt{a}}\right)}\exp(-\mu_n^2 t)$$

$$= \frac{\dfrac{2qa}{L\mu_n^2}\sin\left(\dfrac{\mu_n}{\sqrt{a}}(L-z)\right)}{\sin\left(\dfrac{L\mu_n}{\sqrt{a}}\right)}\exp(-\mu_n^2 t)$$

$$= -(-1)^n\frac{2qa}{L\dfrac{a}{L^2}\left(-\dfrac{\pi}{2}+\pi n\right)^2}$$

$$\times \sin\left(\frac{\left(\dfrac{\pi}{2}+\pi n\right)}{\sqrt{a}}(L-z)\right)\exp(-\mu_n^2 t)$$

$$= \frac{8Lq}{\lambda\pi^2}\frac{\cos\left(\dfrac{z}{L}\dfrac{\pi}{2}(1+2n)\right)}{(1-2n)^2}\exp\left(-(1+2n)^2\mathrm{Fo}\frac{\pi}{4}\right).$$

Here, Fo = $(a/L^2)t$ is the dimensionless Fourier number [7].

The desired solution has the form

$$T_*(t, z) = T_0 + \frac{q}{\lambda}\left((L-x) - \frac{8L}{\pi^2}\sum_{n=0}^{\infty}\frac{\cos\left(\dfrac{z}{L}\dfrac{\pi}{2}(1+2n)\right)}{(1+2n)^2}\right.$$

$$\left.\times\exp\left(-(1+2n)^2\mathrm{Fo}\frac{\pi}{4}\right)\right), \quad z \in [0, L],$$

from whence we can calculate the time to failure τ_T, the depth of heating, and the prefailure temperature profile.

(5) For constant thermal and physical properties of the material, the energy conservation equation in the stationary coordinate system has the form

$$\frac{\partial T(\tau, y)}{\partial\tau} = a\frac{\partial^2 T(\tau, y)}{\partial y^2}. \qquad (9)$$

The thermal balance on the outer surface of the body can be written as

$$q = \left(-\lambda\frac{\partial T}{\partial y}\right)_{y = S(\tau)} + G_{i,\mathrm{w}}\Delta Q, \qquad (10)$$

where $G_{i,\mathrm{w}}$ is the mass rate of evaporation from the surface [1].

It is well known that $G_{i,\mathrm{w}}$ can be determined by the Knudsen–Langmuir formula

$$G_{i,\mathrm{w}} = \frac{a_*(p_i^{\mathrm{s}} - p_i)}{\sqrt{2\pi R T_{\mathrm{w}}/M_i}},$$

where a_* is the accommodation coefficient (for metals, $a_* = 1$), p_i^{s} is the saturated vapor pressure at temperature T_{w}, p_i is the vapor pressure above the surface, M_i is the molecular mass, T_{w} is the temperature of the surface of the sublimating material, and R is the gas constant.

In accordance with the Clausius–Clapeyron equation, the saturated vapor pressure along the phase equilibrium curve is given by

$$p^{\mathrm{s}}(T_{\mathrm{w}}) = \exp(K - \Delta QM/RT_{\mathrm{w}}),$$

where M is the molecular mass and K is an constant found experimentally.

In the stationary frame of reference, the mass flow rate and the coordinate of the outer surface are related as

$$G_{i,\mathrm{w}} = \rho\frac{dS}{d\tau} \quad \text{or} \quad S = \frac{1}{\rho}\int_0^{\tau}G_{i,\mathrm{w}}d\tau. \qquad (11)$$

From the zero time $t = 0$ on, the temperature of the body at a large distance from the outer surface ($y \longrightarrow L$) is assumed to be constant and equal to T_0 (Fig. 2). Let ρv equal $G_{i,\mathrm{w}}$. We introduce a new time frame,

$$t = \tau - \tau_T, \quad 0 \leq t \leq t_k, \qquad (12)$$

where $t = 0$ corresponds to the onset of failure and $t = t_k$ corresponds to the instant when $S(t_k + \tau_T) = L$.

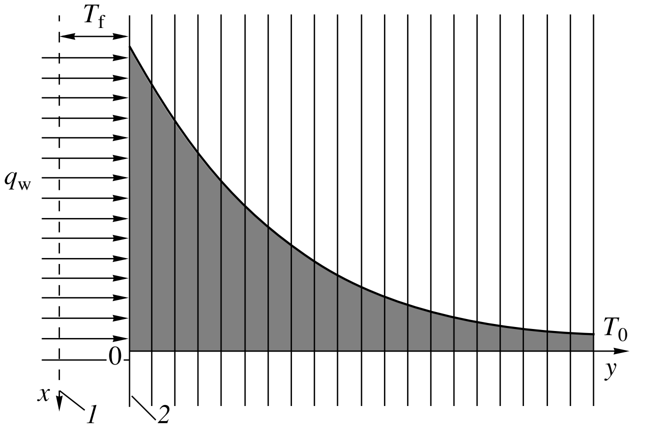


Fig. 2. Model of the melting body with constant failure surface temperature T_{f}: (*1*) initial and (*2*) current position of the failure surface. Zero is the stagnation point.

The need for specifying a boundary condition at the moving surface is eliminated by introducing a dimensionless coordinate and temperature θ,

$$z = y - S(\tau), \quad 0 \le z \le L, \quad \theta = \frac{T - T_0}{T_f - T_0}. \qquad (13)$$

For all t in the range $0 \le t \le t_k$, $z = 0$ corresponds to the failure surface.

With regard to the above substitutions, Eq. (9) assumes the form

$$\frac{\partial\theta}{\partial t} = a\frac{\partial^2\theta}{\partial z^2} + (v + 1)\frac{\partial\theta}{\partial z}; \qquad (14)$$

for $z = 0$, we have $\theta(t, 0) = 1$ and

$$\left.\frac{\partial\theta(t, z)}{\partial z}\right|_{z=0} = \left(\frac{\rho\Delta Q}{\lambda}v - \frac{q}{\lambda}\right)\frac{1}{T_f - T_0}$$
$$= \left(\frac{\Delta Q}{\lambda}\frac{a_*(p_i^s - p_i)}{\sqrt{2\pi R/M_i}} - \frac{q}{\lambda}\right)\frac{1}{T_f - T_0}; \qquad (15)$$

for $z = L$,

$$\theta(t, L) = 0;$$

and for $t = 0$,

$$\theta(0, z) = \frac{T_*(\tau_T, 0) - T_0}{T_p - T_0}, \quad v(0) = 0.$$

In the general case, the vapor pressure at the surface depends on its temperature and, hence, on time. It is easy to check that, for $L \longrightarrow \infty$, linear velocity v of the failure surface increases gradually, tending to steady-state value $v|_{\tau\to\infty} = \bar{v}$. Since this transient lasts indefinitely long, one usually speaks of quasi-stationary, rather than stationary, failure parameters [1].

The quasi-stationary failure conditions, when the temperature profile stops changing with time, can be described by Eqs. (14) and (15) in which $\partial\theta/\partial t \longrightarrow 0$. Then, there exists a solution to the equation

$$\frac{\partial^2\theta}{\partial z^2} + b\frac{\partial\theta}{\partial z} = 0,$$

where

$$b = \frac{1 + \bar{v}}{a},$$

$$\frac{\partial}{\partial z}\left(\frac{\partial\theta}{\partial z} + b\theta\right) = 0, \quad \frac{\partial\theta}{\partial z} + b\theta = C = \text{const},$$

$$\frac{C - C_0\exp(-bx)}{b} = \theta.$$

Substituting boundary conditions (13) and considering that $L \longrightarrow \infty$, we find that

$$\theta = \exp(-bz),$$

$$\left.\frac{\partial\theta(t, z)}{\partial z}\right|_{z=0} = -\frac{(\bar{v} + 1)(T_p - T_0)}{a} = \frac{\rho\Delta Q}{\lambda}\bar{v} - \frac{q}{\lambda},$$

$$\bar{v} = \frac{q + \rho\Delta Q}{\rho(c(T_f - T_0) + \Delta Q)} - 1$$

or

$$\frac{T - T_0}{T_f - T_0} = \exp\left(-\frac{q + \rho\Delta Q}{\rho(c(T_f - T_0) + \Delta Q)}\frac{(y - S(t))}{a}\right).$$

Thus, at $t = 0$, $v(0) = 0$; at $t = \infty$, we obtain

$$v(\infty) = \bar{v} = \frac{q + \rho\Delta Q}{\rho(c(T_f - T_0) + \Delta Q)} - 1,$$

from whence

$$v(t) = \left(\frac{q + \rho\Delta Q}{\rho(c(T_f - T_0) + \Delta Q)} - 1\right)(1 - \exp(-Ct)),$$

where $C > 0$.

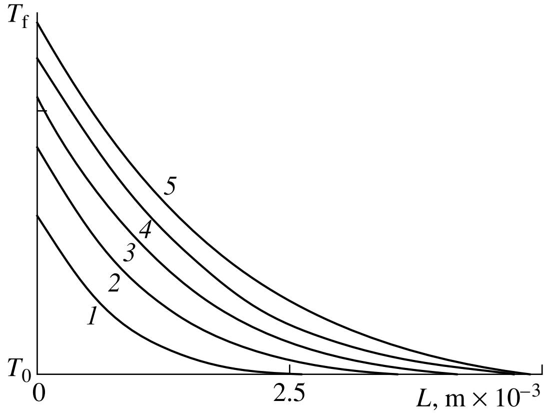


Fig. 3. Temperature distribution across the wall at t = (*1*) 1, (*2*) 5, (*3*) 10, (*4*) 15, and (*5*) 20 s.

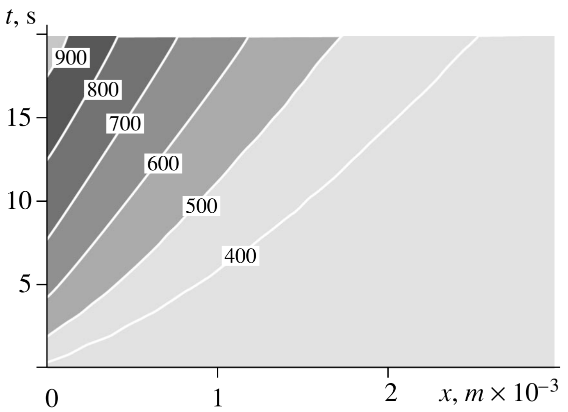


Fig. 4. Temperature distribution across the wall at the time of failure.

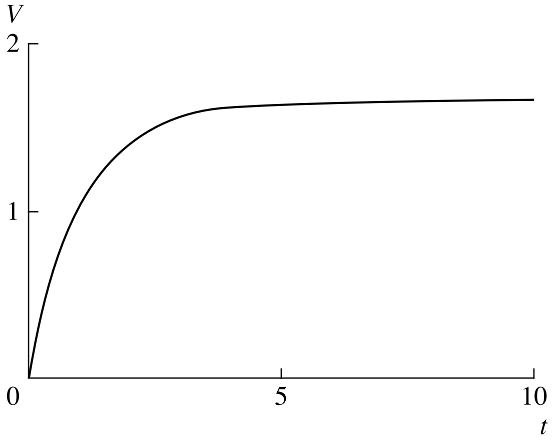


Fig. 5. Saturation of the rate of mass transfer.

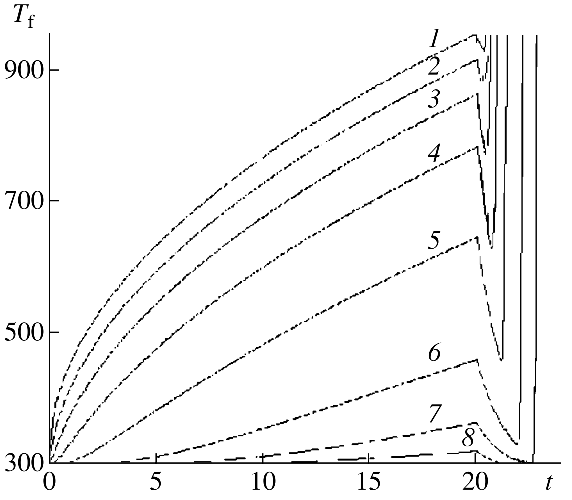


Fig. 6. Temperature profiles at various depths before and at the time of failure. Temperature distributions at depths of (*1*) 0L (over the surface), (*2*) 0.02L, (*3*) 0.05L, (*4*) 0.1L, (*5*) 0.2L, (*6*) 0.4L, (*7*) 0.55L, and (*8*) 0.8L are shown. The failure time $\tau_f = 20$ s.

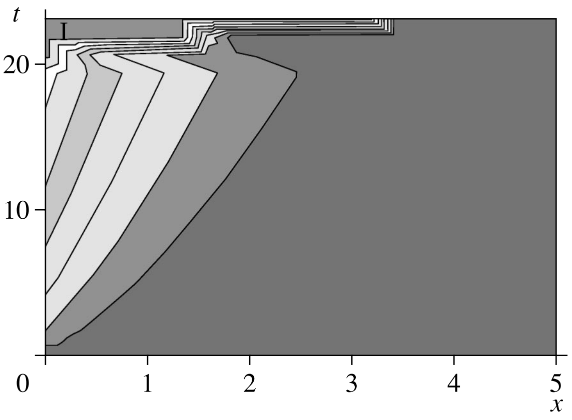


Fig. 7. Depth of wall heating. I, disintegrated part of the wall.

(6) By way of example, we considered a flow about a hemispherically blunted body coated by a heat-reflecting layer of thickness $L = 0.005$ m. The wall parameter at 20°C were $\lambda = 293$ W/(m K), $c = 0.92$ kJ/(kg K), $\rho = 2700$ kg/m^3, and $T_{\text{mel}} = 950$ K.

Figures 3–7 show the results of numerical calculations. It is remarkable that, when the heat flux due to the incoming gas takes a steady-state value, the temperature profile stops varying with time when plotted in the reference frame related to the moving boundary. Even a minor mass removal from the outer surface changes the temperature field configuration.

Figure 3 shows the temperature distribution across the wall at various time instants up to failure. The temperature field across the wall at the time of failure is shown in Fig. 4. As the temperature reaches failure temperature T_f, the heat flux becomes time-independent and linear velocity v_∞ of the failure surface increases gradually, tending toward a constant value (Fig. 5). The wall heats up uniformly until the failure temperature is reached at its surface. In this case, a certain amount of heat is carried away by the material, the temperature field changes, and the temperature decreases slightly; however, heat is supplied again and failure continues (Fig. 6). It can be seen that the rate of wall failure becomes constant with time (Fig. 5). Under the conditions of mass transfer, the depth of heating decreases, since almost all the heat supplied is immediately carried away by the disintegrating coating. In addition, when the material is carried away (i.e., when the temperature of the outer surface is maintained at a constant level), a heat wave propagates across the wall and the temperature profile is described by the same function at any time instant. Figure 7 shows the temperature field distribution during failure. Even if the mass transfer from the surface is minor, the temperature field reconfigures. This effect shows up most vividly at high temperatures. The mass transfer (i.e., the travel of the surface) leads to a situation in which any point, initially being inside the coating, emerges on the outer surface.

From the aforesaid, we can formulate general requirements for heat-reflecting systems based on physicochemical transformations. First, heat-reflecting materials must absorb a large amount of heat during these transformations. Second, they must offer a high heat capacity per unit volume and a high thermal resistance to minimize mechanical mass transfer. Finally, it is desirable that the failure surface have a high temperature and emissivity.

REFERENCES

1. Yu. V. Polezhaev and F. B. Yurevich, *Thermal Protection*, Ed. by A. V. Lykov (Énergiya, Moscow, 1976) [in Russian].

2. N. I. Sidnyaev, in *Proceedings of the 2nd All-Russia Conference on Irreversible Processes in Nature and Technology, Moscow, 2002*, pp. 96–97.
3. N. I. Sidnyaev, V. T. Kalugin, A. I. Lutsenko, *et al.*, in *The Scientific Fundamentals of the Technology in XXI Century*, Ed. by A. I. Leont'ev, N. N. Pilyugin, Yu. V. Polezhaev, *et al.* (UNPTs Énergomash, Moscow, 2000) [in Russian].
4. N. I. Sidnyaev, Vestn. Mosk. Gos. Tekhn. Univ., Ser. Estestvennye Nauki, No. 2, 54 (2001).
5. N. I. Sidnyaev, Mat. Model. **16** (5), 55 (2004).
6. V. S. Zarubin and G. N. Kuvyrkin, *Mathematical Models of Thermal Mechanics* (Fizmatlit, Moscow, 2002) [in Russian].
7. V. S. Zarubin, *Mathematical Modeling in Engineering: A Textbook*, Ed. by V. S. Zarubin and A. P. Krishchenko (Izd. MGTU im. N. E. Baumana, Moscow, 2001) [in Russian].

Translated by N. Wadhwa

Technical Physics, Vol. 50, No. 7, 2005, pp. 835–845. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 19–28.
Original Russian Text Copyright © 2005 by Zharov, Grigor'ev, Shiryaeva.

GASES AND LIQUIDS

Influence of the Viscosity of a Nonlinearly Vibrating Charged Liquid Drop on the Positions of Internal Resonances

A. N. Zharov, A. I. Grigor'ev, and S. O. Shiryaeva
Demidov State University, Sovetskaya ul. 14, Yaroslavl, 150000 Russia
e-mail: grig@uniyar.ac.ru

Received November 10, 2004

Abstract—An asymptotic analytical expression for the generatrix of a viscous charged liquid drop is for the first time derived in the second order of smallness in the axisymmetric initial deformation of the drop. The expression is represented as an infinite series in the roots of the dispersion relation and a finite sum of the numbers of modes specifying the initial deformation. In some of the terms of the analytical expression, the denominators involve the differences between the mode frequencies. These differences may become small under certain values of the charge, causing internal nonlinear resonant mode interaction. Analytical and numerical investigations of the effect of viscosity on the vibrating frequency show that the resonant values of the self-charge of the drop tend to increase with increasing viscosity. The viscosity of the liquid does not affect the spectrum of modes excited via nonlinear mode interaction. © *2005 Pleiades Publishing, Inc.*

(1) Capillary vibration and stability of charged drops are of great scientific and applied interest. This problem has therefore been thoroughly examined in both the linear and nonlinear statements. While in the former case a rigorous solution to the problem considering a viscous drop is easy to find, nonlinear analysis, which became possible about 20 years ago with the advent of advanced computing facilities and dedicated software packages, has been performed for an ideal liquid alone (a total of about 50 publications) (see, e.g., [1] and Refs. cited therein). The most plausible explanation for such a situation is the awkwardness of the mathematics used in calculating nonlinear vibrations of a viscous liquid drop [2]. Nevertheless, taking account of viscosity in such problems brings mathematical models closer to reality and, thus, seems to be of importance.

Below, we report the results of investigation into the effect of viscosity on the positions of internal resonances of a vibrating charged drop. To simplify the mathematics, our consideration is restricted to analysis of an expression for the generatrix of a nonlinearly vibrating viscous liquid drop that is derived for simple forms of initial deformation of the drop (integral terms in this expression are retained) and to analysis of simple (degenerate) internal nonlinear resonances.

(2) Let a spherical drop of an ideal conducting incompressible liquid have radius r_0, density ρ, kinematic viscosity ν, surface tension coefficient σ, and charge Q. In the spherical coordinate system (r, ϑ, φ) with the origin at the center of the drop, the equation for the surface of the drop executing axisymmetric vibrations is written in the form

$$F(r, \vartheta, t) = r - r_0 - \xi(\vartheta, t) = 0, \qquad (1)$$

where t is the time.

The initial deformation of the drop is specified by the relationship

$$t = 0: \xi = \varepsilon \sum_{m \in \Omega} h_m P_m(\mu); \quad \mu \equiv \cos(\vartheta), \qquad (2)$$

where ε is a small parameter proportional to the initial deformation amplitude; $P_m(\mu)$ is the Legendre polynomial of order m; Ω is a set of indices of the modes specifying the initial deformation; and h_m are the constants characterizing the partial contribution of an mth mode to the initial deformation of the drop, $\sum_{m \in \Omega} h_m = O(1)$.

The flow velocity field in the drop is designated as $\mathbf{U}(r, \vartheta, t)$; the pressure field, as $p(r, \vartheta, t)$; and the electric field potentials near the drop and on its surface, as $\phi(r, \vartheta, t)$ and $\phi_s(t)$, respectively.

Mathematically, the problem of capillary vibration of a viscous incompressible conducting charged liquid drop whose shape is defined by (1) and (2) is formulated as [1]

$$\partial_t \mathbf{U} + (\mathbf{U} \cdot \mathbf{\nabla})\mathbf{U} = -\frac{1}{\rho}\operatorname{grad} p + \nu \Delta \mathbf{U}; \quad \operatorname{div} \mathbf{U} = 0;$$

$$\Delta \phi = 0;$$

$$t = 0: \mathbf{U} = 0; \quad r \longrightarrow 0: \mathbf{U} < \infty;$$

$$r \longrightarrow +\infty: \mathbf{\nabla}\phi \longrightarrow 0;$$

$$r = r_0 + \xi(\vartheta, t): \partial_t F + (\mathbf{U} \cdot \mathbf{\nabla})F = 0;$$

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

$$-p + 2\rho\nu \mathbf{n} \cdot (\mathbf{n} \cdot \mathbf{\nabla})\mathbf{U} - \frac{1}{8\pi}(\mathbf{\nabla}\phi)^2 + \sigma(\mathbf{\nabla} \cdot \mathbf{n}) = 0; \qquad (3)$$

$$\phi = \phi_S(t);$$

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

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

$$\int\limits_V r^2 \sin\vartheta dr d\vartheta d\varphi = \frac{4\pi}{3}r_0^3;$$

$$V = \{r, \vartheta, \varphi | 0 \le r \le r_0 + \xi;\ 0 \le \vartheta \le \pi;\ 0 \le \varphi \le 2\pi\};$$

$$\int\limits_V \mathbf{r} r^2 \sin\vartheta dr d\vartheta d\varphi = 0,$$

where symbol ∂_t means the partial derivative with respect to t and $\mathbf{n}$ and $\boldsymbol{\tau}$ are the unit vectors of the normal to the tangent, and Δ is the Laplacian.

The above set of equations is nonlinear; therefore, solving it by the method of direct expansion [3] requires that all desired quantities be represented as expansions in small parameter ε,

$$\xi(\vartheta, t) = \varepsilon\xi^{(1)}(\vartheta, t) + \varepsilon^2\xi^{(2)}(\vartheta, t) + O(\varepsilon^3);$$

$$\mathbf{U}(r, \vartheta, t) = \varepsilon U_r^{(1)}(r, \vartheta, t)\mathbf{e}_r + \varepsilon^2 U_r^{(2)}(r, \vartheta, t)\mathbf{e}_r$$

$$+ \varepsilon U_\vartheta^{(1)}(r, \vartheta, t)\mathbf{e}_\vartheta + \varepsilon^2 U_\vartheta^{(2)}(r, \vartheta, t)\mathbf{e}_\vartheta + O(\varepsilon^3);$$

$$p(r, \vartheta, t) = p^{(0)}(r, \vartheta, t) + \varepsilon p^{(1)}(r, \vartheta, t)$$

$$+ \varepsilon^2 p^{(2)}(r, \vartheta, t) + O(\varepsilon^3); \tag{4}$$

$$\phi(r, \vartheta, t) = \phi^{(0)}(r, t) + \varepsilon\phi^{(1)}(r, \vartheta, t)$$

$$+ \varepsilon^2\phi^{(2)}(r, \vartheta, t) + O(\varepsilon^3);$$

$$\phi_s(t) = \phi_s^{(0)}(t) + \varepsilon\phi_s^{(1)}(t) + \varepsilon^2\phi_s^{(2)}(t) + O(\varepsilon^3),$$

where $\mathbf{e}_r$ and $\mathbf{e}_V$ are the unit vectors of the spherical coordinate system.

(3) Substituting expansions (4) into set (3) and equating the coefficients multiplying terms with different powers of the small parameter, we formulate the problem in the zeroth, first, and second orders of smallness.

(i) In the zeroth order of smallness, we have $\Delta\phi^{(0)} = 0$.

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

$$r = r_0: \phi^{(0)} = \phi_s^{(0)}(t), \quad \int\limits_{-1}^{1} r_0^2\partial_r\phi^{(0)} d(\cos\vartheta) = -2Q,$$

$$p^{(0)} + \frac{Q^2}{8\pi r_0^4} = \frac{2\sigma}{r_0}.$$

Solving this problem yields

$$\phi^{(0)} = \frac{Q}{r}; \quad \phi_s^{(0)} = \frac{Q}{r_0}. \tag{5}$$

(ii) In the first order of smallness, we get

$$\partial_t U_r^{(1)} = -\frac{1}{\rho}\partial_r p^{(1)} + \nu\left(\frac{1}{r^2}\partial_{\vartheta\vartheta}U_r^{(1)} + \frac{\cot(\vartheta)}{r^2}\partial_\vartheta U_r^{(1)}\right.$$

$$\left. - \frac{1}{r}\partial_{r\vartheta}U_\vartheta^{(1)} - \frac{\cot(\vartheta)}{r}\partial_r U_\vartheta^{(1)} - \frac{1}{r^2}\partial_\vartheta U_\vartheta^{(1)} - \frac{\cot(\vartheta)}{r^2}U_\vartheta^{(1)}\right);$$

$$\partial_t U_\vartheta^{(1)} = -\frac{1}{\rho}\frac{1}{r}\partial_\vartheta p^{(1)} + \nu\left(\partial_{rr}U_\vartheta^{(1)} + \frac{2}{r}\partial_r U_\vartheta^{(1)} - \frac{1}{r}\partial_{r\vartheta}U_r^{(1)}\right);$$

$$\partial_r U_r^{(1)} + \frac{2}{r}U_r^{(1)} + \frac{1}{r}\partial_\vartheta U_\vartheta^{(1)} + \frac{\cot(\vartheta)}{r}U_\vartheta^{(1)} = 0; \ \Delta\phi^{(1)} = 0;$$

$$t = 0: \mathbf{U}^{(1)} = 0; \quad \xi^{(1)} = \varepsilon\sum_{m\in\Omega} h_m P_m(\mu);$$

$$r \longrightarrow 0: \mathbf{U}^{(1)} < \infty; \quad r \longrightarrow +\infty: \nabla\phi^{(1)} \longrightarrow 0;$$

$$r = r_0: \phi^{(1)} + \xi^{(1)}\partial_r\phi^{(0)} = \phi_s^{(1)}(t); \quad \partial_t\xi^{(1)} = U_r^{(1)}; \tag{6}$$

$$\partial_r U_\vartheta^{(1)} + \frac{1}{r}\partial_\vartheta U_r^{(1)} - \frac{1}{r}U_\vartheta^{(1)} = 0;$$

$$-p^{(1)} + 2\rho\nu\partial_r U_r^{(1)} - \frac{1}{4\pi}\partial_r\phi^{(0)}(\partial_r\phi^{(1)} + \xi^{(1)}\partial_{rr}\phi^{(0)})$$

$$-\frac{\sigma}{r_0^2}(2 + \Delta_\Omega)\xi^{(1)} = 0;$$

$$\int\limits_{-1}^{1}(r_0\partial_r\phi^{(1)} + \xi^{(1)}(r_0\partial_{rr}\phi^{(0)} + 2\partial_r\phi^{(0)}))d(\mu) = 0;$$

$$\int\limits_{-1}^{1}\xi^{(1)}d(\mu) = 0; \quad \int\limits_{-1}^{1}\xi^{(1)}P_1(\mu)d(\mu) = 0,$$

where Δ_Ω is the angular part of the Laplacian.

A solution to problem (6) is easy to find in the form [1]

$$\xi^{(1)}(\vartheta, t) = \sum_{n\in\Omega}\xi_n^{(1)}(t)P_n(\mu);$$

$$U_r^{(1)}(r, \vartheta, t) = \sum_{n\in\Omega}U_{rn}^{(1)}(r, t)P_n(\mu); \tag{7}$$

$$U_\vartheta^{(1)}(r, \vartheta, t) = \sum_{n\in\Omega}U_{\vartheta n}^{(1)}(r, t)\partial_\vartheta P_n(\mu);$$

$$p_n^{(1)}(r,\vartheta,t) = \sum_{n\in\Omega} p_n^{(1)}(r,t)P_n(\mu),$$

where

$$\xi_n^{(1)}(t) = \sum_{j=1}^{+\infty} a_{\xi_n}(S_n^{(j)})\exp(S_n^{(j)}t);$$

$$U_{rn}^{(1)}(r,t) = \sum_{j=1}^{+\infty}\left(a_n(S_n^{(j)})\left(\frac{r}{r_0}\right)^{n-1}\right.$$

$$\left.+\,b_n(S_n^{(j)})\frac{1}{r}\frac{j_n(\chi_n^{(j)}r)}{j_n(\chi_n^{(j)}r_0)}\right)\exp(S_n^{(j)}t);$$

$$U_{\vartheta n}^{(1)}(r,t) = \sum_{j=1}^{+\infty}\left(a_n(S_n^{(j)})\left(\frac{r}{r_0}\right)^{n-1} + b_n(S_n^{(j)})\left(\frac{1}{r}\frac{j_n(\chi_n^{(j)}r)}{j_n(\chi_n^{(j)}r_0)}\right.\right.$$

$$\left.\left.+\,\frac{\chi_n^{(j)}}{n+1}\frac{j_{n+1}(\chi_n^{(j)}r)}{j_n(\chi_n^{(j)}r_0)}\right)\right)\frac{\exp(S_n^{(j)}t)}{n};$$

$$a_{\xi_n}(S_n^{(j)}) = \left(S_n^{(j)} + 2(n-1)(2n+1)\frac{\nu}{r_0^2} + 2(n-1)^2(n+1)\right.$$

$$\left.\times\frac{\nu}{\eta_n^{(0)}(\chi_n^{(j)})r_0^2}\right)\frac{h_n}{\partial_{S_n^{(j)}}D_n(S_n^{(j)})};$$

$$a_n(S_n^{(j)}) = \left((2(n^2-1)+(r_0\chi_n^{(j)})^2)\right.$$

$$\left.\times\frac{1}{2\chi_n^{(j)}r_0}\frac{j_n(\chi_n^{(j)}r_0)}{j_{n+1}(\chi_n^{(j)}r_0)} - 1\right)\frac{h_n}{\eta_n^0(\chi_n^{(j)})}\frac{\omega_n^2}{\partial_{S_n^{(j)}}D_n(S_n^{(j)})};$$

$$b_n(S_n^{(j)}) = 2(n^2-1)\left(1-\frac{2}{\chi_n^{(j)}r_0}\frac{j_{n+1}(\chi_n^{(j)}r_0)}{j_n(\chi_n^{(j)}r_0)}\right)^{-1}$$

$$\times\frac{h_n\omega_n^2\nu}{r_0S_n^{(j)}\partial_{S_n^{(j)}}D_n(S_n^{(j)})};\quad \chi_n^{(j)} = \sqrt{\frac{S_n^{(j)}}{\nu}};$$

$$\partial_{S_n^{(j)}}D_n(S_n^{(j)}) = 2S_n^{(j)} + 2(n-1)(2n+1)\frac{\nu}{r_0^2}$$

$$+\,(n-1)^2(n+1)\frac{\nu}{r_0^2}\left(2+\frac{(2n+1)\chi_n^{(j)}r_0}{2}\frac{j_n(\chi_n^{(j)}r_0)}{j_{n+1}(\chi_n^{(j)}r_0)}\right.$$

$$\left.+\,\frac{(\chi_n^{(j)}r_0)^2}{2}\left(1-\left(\frac{j_n(\chi_n^{(j)}r_0)}{j_{n+1}(\chi_n^{(j)}r_0)}\right)^2\right)\right)\frac{1}{\eta_n^0(\chi_n^{(j)})};$$

$$D_n(S_n^{(j)}) = (S_n^{(j)})^2 + 2(n-1)(2n+1)\frac{S_n^{(j)}\nu}{r_0^2}$$

$$+\,2(n-1)^2(n+1)\frac{S_n^{(j)}\nu}{r_0^2}\left(1-\frac{\chi_n^{(j)}r_0}{2}\frac{j_n(\chi_n^{(j)}r_0)}{j_{n+1}(\chi_n^{(j)}r_0)}\right)^{-1} + \omega_n^2;$$

$$\omega_n^2 = \frac{\sigma}{\rho r_0^3}n(n-1)(n+2-W);\quad W = \frac{Q^2}{4\pi\sigma r_0^3};$$

$$\eta_n^0(\chi_n^{(j)}) = 1-\frac{r_0\chi_n^{(j)}}{2}\frac{j_n(\chi_n^{(j)}r_0)}{j_{n+1}(\chi_n^{(j)}r_0)}.$$

Here, $S_n^{(j)}$ is the root of the dispersion relation $D_n(S_n^{(j)}) = 0$ and $j_n(\chi_n^{(j)}r_0)$ is the modified spherical Bessel function of $\chi_n^{(j)}r_0$ of the first kind [4].

(iii) To find the components of a second-order solution, we write

$$\partial_t U_r^{(2)} + U_r^{(1)}\partial_r U_r^{(1)} + \frac{1}{r}U_\vartheta^{(1)}\partial_\vartheta U_r^{(1)} - \frac{1}{r}(U_\vartheta^{(1)})^2$$

$$= -\frac{1}{\rho}\partial_r p^{(2)} + \nu\left(\frac{1}{r^2}\partial_{\vartheta\vartheta}U_r^{(2)} + \frac{\cot(\vartheta)}{r^2}\partial_\vartheta U_r^{(2)} - \frac{1}{r}\partial_{r\vartheta}U_\vartheta^{(2)}\right.$$

$$\left.-\,\frac{\cot(\vartheta)}{r}\partial_r U_\vartheta^{(2)} - \frac{1}{r^2}\partial_\vartheta U_\vartheta^{(2)} - \frac{\cot(\vartheta)}{r^2}U_\vartheta^{(2)}\right);$$

$$\partial_t U_\vartheta^{(2)} + U_r^{(1)}\partial_r U_\vartheta^{(1)} + \frac{1}{r}U_\vartheta^{(1)}\partial_\vartheta U_\vartheta^{(1)} + \frac{1}{r}U_r^{(1)}U_\vartheta^{(1)}$$

$$= -\frac{1}{\rho}\frac{1}{r}\partial_\vartheta p^{(2)} + \nu\left(\partial_{rr}U_\vartheta^{(2)} + \frac{2}{r}\partial_r U_\vartheta^{(2)} - \frac{1}{r}\partial_{r\vartheta}U_r^{(2)}\right);$$

$$\partial_t U_r^{(2)} + \frac{2}{r}U_r^{(2)} + \frac{1}{r}\partial_\vartheta U_\vartheta^{(2)} + \frac{\cot(\vartheta)}{r}U_\vartheta^{(2)} = 0;$$

$$\Delta\phi^{(2)} = 0;$$

$$t = 0:\ \mathbf{U}^{(2)} = 0;\quad \xi^{(2)} = -\frac{1}{r_0}\sum_{m\in\Omega}\frac{h_m^2}{2m+1}P_0(\mu)$$

$$-\frac{9}{r_0}\sum_{m\in\Omega}\frac{(m+1)h_m h_{m+1}}{(2m+1)(2m+3)}P_1(\mu);$$

$$r \longrightarrow 0:\ \mathbf{U}^{(2)} < \infty;\quad r \longrightarrow +\infty:\ \boldsymbol{\nabla}\phi^{(2)} \longrightarrow 0;$$

$$r = r_0:$$

$$\phi^{(2)} + \xi^{(2)}\partial_r\phi^{(0)} + \frac{1}{2}(\xi^{(1)})^2\partial_{rr}\phi^{(0)} + \xi^{(1)}\partial_r\phi^{(1)} = \phi_s^2(t);$$

$$
\begin{gathered}
\int_{-1}^{1}\Big[r_0^2\partial_r\phi^{(2)} + r_0\xi^{(1)}(r_0\partial_{rr}\phi^{(1)} + 2\partial_r\phi^{(1)}) + r_0\xi^{(2)}(r_0\partial_{rr}\phi^{(0)} \\
+ 2\partial_r\phi^{(0)}) + (\xi^{(1)})^2\left(\frac{1}{2}r_0^2\partial_{rrr}\phi^{(0)} + 2r_0\partial_{rr}\phi^{(0)} + \partial_r\phi^{(0)}\right) \\
- \partial_\vartheta\xi^{(1)}\partial_\vartheta\phi^{(1)}\Big]d(\mu) = 0; \\
\int_{-1}^{1}(r_0\xi^{(2)} + (\xi^{(1)})^2)d(\mu) = 0; \\
\int_{-1}^{1}(2r_0\xi^{(2)} + 3(\xi^{(1)})^2)P_1(\mu)d(\mu) = 0; \\
-\partial_t\xi^{(2)}(\vartheta, t) + U_r^{(2)} + \partial_r U_r^{(1)}\xi^{(1)}(\vartheta, t) \\
- \frac{1}{r_0}U_\vartheta^{(1)}\partial_\vartheta\xi^{(1)}(\vartheta, t) = 0; \\
\frac{1}{r_0}\partial_\vartheta U_r^{(2)} + \partial_r U_\vartheta^{(2)} - \frac{1}{r_0}U_\vartheta^{(2)} + \left(\frac{1}{r_0}\partial_{r\vartheta}U_r^{(1)} - \frac{1}{r_0^2}\partial_\vartheta U_r^{(1)}\right. \\
\left. + \partial_{rr}U_\vartheta^{(1)} - \frac{1}{r_0}\partial_r U_\vartheta^{(1)} + \frac{1}{r_0^2}U_\vartheta^{(1)}\right)\xi^{(1)}(\vartheta, t) \\
- 2\left(\frac{1}{r_0^2}\partial_\vartheta U_\vartheta^{(1)} + \frac{1}{r_0^2}U_r^{(1)} - \frac{1}{r_0}\partial_r U_r^{(1)}\right)\partial_\vartheta\xi^{(1)}(\vartheta, t) = 0; \\
-p^{(2)} - \frac{\sigma}{r_0^2}(2 + \Delta_\Omega)\xi^{(2)} + \frac{2\sigma}{r_0^3}\xi^{(1)}(1 + \Delta_\Omega)\xi^{(1)} \\
- \frac{1}{8\pi}[2\xi^{(2)}\partial_{rr}\phi^{(0)}\partial_r\phi^{(0)} + (\xi^{(1)})^2((\partial_{rr}\phi^{(0)})^2 \\
+ \partial_{rrr}\phi^{(0)}\partial_r\phi^{(0)}) + \frac{1}{r_0^2}(\partial_\vartheta\phi^{(1)})^2 + (\partial_r\phi^{(1)})^2 + 2\partial_r\phi^{(2)}\partial_r\phi^{(0)} \\
+ 2\xi^{(1)}(\partial_{rr}\phi^{(0)}\partial_r\phi^{(1)} + \partial_{rr}\phi^{(1)}\partial_r\phi^{(0)})] + 2\rho\nu\partial_r U_r^{(2)} \\
- (\partial_r p^{(1)} - 2\rho\nu\partial_{rr}U_r^{(1)})\xi^{(1)}(\vartheta, t) - 2\rho\nu\left(\frac{1}{r_0^2}\partial_\vartheta U_r^{(1)}\right. \\
\left. + \frac{1}{r_0}\partial_r U_\vartheta^{(1)} - \frac{1}{r_0^2}U_\vartheta^{(1)}\right)\partial_\vartheta\xi^{(1)}(\vartheta, t) = 0.
\end{gathered}
\tag{8}
$$

Substituting the quantities of the zeroth, (5), and first, (7), orders of smallness into (8) gives a set of inhomogeneous partial differential equations, which can be solved using the Laplace transformation

$$
f(S) = \int_0^{+\infty} f(t)\exp(-St)dt = \mathcal{F}[f(t)].
$$

The Laplace transforms of the second-order quantities are now expanded in an infinite set of Legendre polynomials or their derivatives,

$$
\begin{gathered}
\xi^{(2)}(\vartheta, S) = \sum_{n=0}^{+\infty}\xi_n^{(2)}(S)P_n(\mu); \\
\phi^{(2)}(r, \vartheta, S) = \sum_{n=0}^{+\infty}\phi_n^{(2)}(r, S)P_n(\mu); \\
U_r^{(2)}(r, \vartheta, S) = \sum_{n=0}^{+\infty}U_{rn}^{(2)}(r, S)P_n(\mu); \\
U_\vartheta^{(2)}(r, \vartheta, S) = \sum_{n=0}^{+\infty}U_{\vartheta n}^{(2)}(r, S)\partial_\vartheta P_n(\mu); \\
p^{(2)}(r, \vartheta, S) = \sum_{n=0}^{+\infty}p_n^{(2)}(r, S)P_n(\mu).
\end{gathered}
\tag{9}
$$

Substituting expansions (5), (7), and (9) into (8) yields a fourth-order inhomogeneous differential equation for $U_{rn}^{(2)}(r, S)$,

$$
\begin{gathered}
\left(\partial_{rr} + \frac{4}{r}\partial_r - \frac{(n-1)(n+2)}{r^2}\right) \\
\times\left(\partial_{rr} + \frac{4}{r}\partial_r - \frac{(n-1)(n+2)}{r^2} - \frac{S}{\nu}\right)U_{rn}^{(2)}(r, S) \\
= \frac{n(n+1)}{\nu}\sum_{k, m \in \Omega}\sum_{l, g = 1}^{+\infty}\frac{f_{kmn}^{lg}(r, S_k^{(l)}, S_m^{(g)})}{S - S_k^{(l)} - S_m^{(g)}}.
\end{gathered}
\tag{10}
$$

Here,

$$
\begin{gathered}
f_{kmn}^{lg}(r, S_k^{(l)}, S_m^{(g)}) = f_{kmn}^1 a_k(S_k^{(l)})a_m(S_m^{(g)})r^{m+k-5}/r_0^{m+k-2} \\
+ (f_{kmn}^2(\chi_m^{(g)}r)^2 + f_{kmn}^3)a_k(S_k^{(l)})b_m(S_m^{(g)})r^{k-5} \\
\times j_m(\chi_m^{(g)}r)/(r_0^{k-1}j_m(\chi_m^{(g)}r_0)) + (f_{kmn}^4(\chi_m^{(g)}r)^2 + f_{kmn}^5) \\
\times a_k(S_k^{(l)})b_m(S_m^{(g)})\chi_m^{(g)}r^{k-4}j_{m+1}(\chi_m^{(g)}r)/(r_0^{k-1}j_m(\chi_m^{(g)}r_0)) \\
+ (f_{kmn}^2(\chi_m^{(g)}r)^2 + f_{kmn}^1)b_k(S_k^{(l)})b_m(S_m^{(g)})j_k(\chi_k^{(l)}r) \\
\times j_m(\chi_m^{(g)}r)/(r^5 j_k(\chi_k^{(l)}r_0)j_m(\chi_m^{(g)}r_0)) + (f_{kmn}^4(\chi_m^{(g)}r)^2
\end{gathered}
$$

$$+ f^6_{kmn}(\chi^{(l)}_k r)^2 + f^5_{kmn})b_k(S^{(l)}_k)b_m(S^{(g)}_m)\chi^{(g)}_m j_k(\chi^{(l)}_k r)$$
$$\times j_{m+1}(\chi^{(g)}_m r)/(r^4 j_k(\chi^{(l)}_k r_0)j_m(\chi^{(g)}_m r_0)) + f^7_{kmn}b_k(S^{(l)}_k)$$
$$\times b_m(S^{(g)}_m)\chi^{(l)}_k\chi^{(g)}_m j_{k+1}(\chi^{(l)}_k r)$$
$$\times j_{m+1}(\chi^{(g)}_m r)/(r^3 j_k(\chi^{(l)}_k r_0)j_m(\chi^{(g)}_m r_0));$$

$$f^1_{kmn} = (m-1)((m^2 + k(m+2) - 2)\Gamma_{kmn}/(m(m+1)) + 2\Gamma_{mkn}/(k(k+1)) + ((k(m+4) + m - 2)\Lambda_{kmn} + (k(m-2) + m + 4)\Lambda_{mkn})/(k(k+1)m(m+1)) - K_{kmn} - (k^2 + 2m + 1)\alpha_{kmn}/(k(k+1)m(m+1)));$$

$$f^4_{kmn} = \Gamma_{kmn}/(m(m+1));$$

$$f^2_{kmn} = (m+k)\Gamma_{kmn}/(m(m+1)) + (\Lambda_{kmn} + \Lambda_{mkn})/(km(m+1));$$

$$f^3_{kmn} = (k+m-2)(\Gamma_{kmn} + \Gamma_{mkn} - K_{kmn} + (\Lambda_{kmn} + \Lambda_{mkn} - \alpha_{kmn})/(km));$$

$$f^5_{kmn} = \Gamma_{kmn} + \Gamma_{mkn} - K_{kmn} + (k-2)(\Lambda_{kmn} + \Lambda_{mkn} - \alpha_{kmn})/(km(m+1));$$

$$f^6_{kmn} = \Gamma_{mkn}/(k(k+1)) + (\Lambda_{kmn} + \Lambda_{mkn})/(k(k+1)m(m+1));$$

$$\chi^{(l)}_k = \sqrt{S^{(l)}_k/\nu};$$

$$f^7_{kmn} = (2\Lambda_{kmn} - 4\Lambda_{mkn} + \alpha_{kmn})/(k(k+1)m(m+1));$$

$$\chi^{(g)}_m = \sqrt{S^{(g)}_m/\nu};$$

$$\alpha_{kmn} = -C^{n0}_{k0m0}C^{n0}_{k(-1)m1}\sqrt{k(k+1)m(m+1)};$$

$$\Gamma_{kmn} = (2n+1)\alpha_{nmk}/(n(n+1)(2k+1));$$

$$K_{kmn} = (C^{n0}_{k0m0})^2; \quad \Lambda_{kmn} = (2n+1)\left(\sum_{j=1}^{[m/2]}\alpha_{n,k,m-2j} - m^2\alpha_{nkm}/(2m+1)\right)/(n(n+1));$$

and C^{n0}_{k0m0} and $C^{n0}_{k(-1)m1}$ are Clebsch–Gordan coefficients.

At $r \longrightarrow 0$, Eq. (10) has a limited solution of type

$$U^{(2)}_{rn}(r,S) = A_n(S)r^{n-1} + B_n(S)\frac{1}{r}j_n(\chi r)$$
$$+ \sum_{k,m\in\Omega}\sum_{l,g=1}^{+\infty}\frac{n(n+1)}{S - S^{(l)}_k - S^{(g)}_m}\int_0^r\left\{\frac{1}{(2n+1)}\frac{1}{S}\left(\frac{\tau^{n+3}}{r^{n+2}} - \frac{r^{n-1}}{\tau^{n-2}}\right) + (-1)^n\frac{1}{\sqrt{S}\sqrt{\nu}}\frac{\tau^3}{r}(j_n(\chi r)y_n(\chi\tau) - y_n(\chi r)j_n(\chi\tau))\right\} \tag{11}$$
$$\times f^{lg}_{kmn}(\tau, S^{(l)}_k, S^{(g)}_m)d\tau,$$

where $y_n(\chi r)$ is the nth-order spherical modified Bessel function of the second kind [4], $A_n(S)$ and $B_n(S)$ are arbitrary constants, and $\chi = \sqrt{S/\nu}$.

Substituting (11) into the boundary conditions, we arrive at a set of three linear algebraic equations in $A_n(S)$, $B_n(S)$, and $\xi^{(2)}_n(S)$, which is very awkward. In view of this circumstance and the goal of our investigation, we present only the expression for $\xi^{(2)}_n(S)$,

$$\xi^{(2)}_0(S) = -\frac{1}{r_0}\sum_{m\in\Omega}\frac{1}{2m+1}\mathcal{F}[(\xi^{(1)}_m(t))^2];$$

$$\xi^{(2)}_1(S) = -\frac{9}{r_0}\sum_{m\in\Omega}\frac{(m+1)}{(2m+1)(2m+3)}\mathcal{F}[\xi^{(1)}_m(t)\xi^{(1)}_{m+1}(t)]; \tag{12}$$

$$\xi^{(2)}_n(S) = \sum_{k,m\in\Omega}\sum_{l,g=1}^{+\infty}\frac{\zeta^{lg}_{kmn}(S, S^{(l)}_k, S^{(g)}_m)}{D_n(S)(S - S^{(l)}_k - S^{(g)}_m)}; \quad n \ge 2.$$

Here,

$$\zeta^{lg}_{kmn}(S, S^{(l)}_k, S^{(g)}_m) = \left((\chi r_0)^2 + (n-1)(3n+1) + \frac{(n^2-1)}{\eta^0_n(\chi)}\left(2n - 1 - \frac{r_0\chi}{2}\frac{j_n(\chi r_0)}{j_{n+1}(\chi r_0)}\right)\right)$$
$$\times\frac{2n(n+1)}{2n+1}\frac{r_0^{n-3}}{\chi^2}\int_0^{r_0}\frac{f^{lg}_{kmn}(r, S^{(l)}_k, S^{(g)}_m)}{r^{n-2}}dr$$
$$-\left((\chi r_0)^2 + 2(n+1)(n(n-3) - 1) + \frac{2(n+1)}{\eta^0_n(\chi)}\times\left(n(2n^2 + n - 1) + 1 - n(n+2)\frac{r_0\chi}{2}\frac{j_n(\chi r_0)}{j_{n+1}(\chi r_0)}\right)\right)$$
$$\times\frac{n}{2n+1}\frac{1}{\chi^2}\frac{1}{r_0^{n+4}}\int_0^{r_0}r^{n+3}f^{lg}_{kmn}(r, S^{(l)}_k, S^{(g)}_m)dr$$

$$-n(n+1)\frac{j_n(\chi r_0)}{\eta_n^0(\chi)}\bigg(2(\chi r_0)^2+4n(n-1)(n+2)$$

$$-\chi r_0\bigg((\chi r_0)^2+2(n-1)(2n+1)\frac{j_n(\chi r_0)}{j_{n+1}(\chi r_0)}\bigg)\bigg)\frac{(-1)^n}{\chi r_0^3}$$

$$\times\int_0^{r_0} r^3 y_n(\chi r)\,f_{kmn}^{lg}(r,S_k^{(l)},S_m^{(g)})dr+\frac{n(n+1)}{\eta_n^0(\chi)}$$

$$\times\Bigg(\bigg(\bigg((n+1)(\chi r_0)^2+4n(n-1)(n+2)$$

$$-\chi r_0((\chi r_0)^2+2(n-1)(2n+1))\frac{j_n(\chi r_0)}{j_{n+1}(\chi r_0)}\bigg)y_n(\chi r_0)$$

$$-(n-1)(\chi r_0)^2\frac{j_n(\chi r_0)}{j_{n+1}(\chi r_0)}y_{n+1}(\chi r_0)\bigg)\frac{(-1)^n}{\chi r_0^3}$$

$$\times\int_0^{r_0} r^3 j_n(\chi r)\,f_{kmn}^{lg}(r,S_k^{(l)},S_m^{(g)})dr+\bigg(\frac{\eta_n^1(\chi)}{\eta_n^0(\chi)}n(n+1)$$

$$\times\bigg(\frac{2\Lambda_{kmn}}{m}-\frac{2(m-2)(m-1)\Gamma_{kmn}}{m}-2(m-2)\Gamma_{mkn}\bigg)$$

$$-\frac{(D_n(S)-\omega_n^2)r_0^2}{\nu^2\chi^2}\bigg(\frac{\alpha_{kmn}}{m}-(m-1)K_{kmn}\bigg)\bigg)\frac{\nu}{r_0^4}a_{\xi k}(S_k^{(l)})$$

$$\times(a_m(S_m^{(g)})r_0+b_m(S_m^{(g)}))+\bigg(\frac{4(m-1)\alpha_{kmn}}{m}$$

$$-((\chi_m^{(g)}r_0)^2+2(m-2)(m-1))K_{kmn}\bigg)\frac{n\nu}{r_0^3}a_{\xi k}(S_k^{(l)})a_m(S_m^{(g)})$$

$$+\Bigg(((\chi_m^{(g)}r_0)^2+2(m^2-1))\frac{2\alpha_{kmn}}{m(m+1)}$$

$$-((\chi_m^{(g)}r_0)^2+(m-2)(m-1))2K_{kmn}$$

$$-\frac{\eta_n^1(\chi)}{\eta_n^0(\chi)}\frac{(n+1)(m-2)}{m(m+1)}\Bigg)(\chi_m^{(g)}r_0)^2\Gamma_{kmn}\Bigg)\frac{n\nu}{r_0^4}a_{\xi k}(S_k^{(l)})b_m(S_m^{(g)})$$

$$+\frac{n\sigma}{2r_0^4\rho}((4(k^2+k-1)+(3+k(m+1)$$

$$+m(2n-2m-7))W)K_{kmn}+W\alpha_{kmn})a_{\xi k}(S_k^{(l)})a_{\xi m}(S_m^{(g)})$$

$$+\Bigg(\frac{\eta_n^1(\chi)}{\eta_n^0(\chi)}\bigg(\frac{2\Lambda_{kmn}}{m(m+1)}-((\chi_m^{(g)}r_0)^2+2(m^2+m+2))$$

$$\times\frac{\Gamma_{kmn}}{m(m+1)}-2\Gamma_{kmn}\bigg)n(n+1)$$

$$-\frac{(D_n(S)-\omega_n^2)r_0^2}{S\nu}\bigg(\frac{\alpha_{kmn}}{m(m+1)}-K_{kmn}\bigg)+8nK_{kmn}$$

$$-\frac{4n\alpha_{kmn}}{m(m+1)}\Bigg)a_{\xi k}(S_k^{(l)})b_m(S_m^{(g)})\chi_m^{(g)}\frac{\nu}{r_0^3}\frac{j_{m+1}(\chi_m^{(g)}r_0)}{j_m(\chi_m^{(g)}r_0)}$$

$$-\bigg(\Gamma_{kmn}+\frac{(km-k+m+3)\Lambda_{kmn}}{k(k+1)m(m+1)}+\frac{2(m-1)\Lambda_{mkn}}{k(k+1)m(m+1)}\bigg)$$

$$\times\frac{n}{r_0}a_k(S_k^{(l)})a_m(S_m^{(g)})-\bigg(\Gamma_{kmn}+\Gamma_{mkn}+\frac{\Gamma_{kmn}}{m(m+1)}$$

$$\times(\chi_m^{(g)}r_0)^2+\frac{\Lambda_{kmn}+\Lambda_{mkn}}{km}\bigg)\frac{n}{r_0^2}a_k(S_k^{(l)})b_m(S_m^{(g)})$$

$$-\bigg(\Gamma_{kmn}+\frac{\Gamma_{kmn}}{m(m+1)}(\chi_m^{(g)}r_0)^2+\frac{(km-k+m+3)\Lambda_{kmn}}{k(k+1)m(m+1)}$$

$$+\frac{2(m-1)\Lambda_{mkn}}{k(k+1)m(m+1)}\bigg)\frac{n}{r_0^3}b_k(S_k^{(l)})b_m(S_m^{(g)})$$

$$-\frac{n}{r_0^2}\bigg(\frac{\Lambda_{kmn}+\Lambda_{mkn}}{km(m+1)}\bigg)\chi_m^{(g)}(r_0a_k(S_k^{(l)})+b_k(S_k^{(l)}))b_m(S_m^{(g)})$$

$$\times\frac{j_{m+1}(\chi_m^{(g)}r_0)}{j_m(\chi_m^{(g)}r_0)}-\frac{n}{r_0^2}\frac{\Lambda_{kmn}}{k(k+1)m(m+1)}\chi_k^{(l)}\chi_m^{(g)}b_k(S_k^{(l)})$$

$$\times b_m(S_m^{(g)})\frac{j_{k+1}(\chi_k^{(l)}r_0)}{j_k(\chi_k^{(l)}r_0)}\frac{j_{m+1}(\chi_m^{(g)}r_0)}{j_m(\chi_m^{(g)}r_0)};$$

$$\eta_n^1(\chi) = n-\frac{r_0\chi}{2}\frac{j_n(\chi r_0)}{j_{n+1}(\chi r_0)}.$$

(4) The form of expression (12) implies that it has an infinite countable number of singular points that are defined by the conditions $D_n(S) = 0$ and $S - S_k^{(l)} - S_m^{(g)} = 0$ and are simple poles. In addition, expression (12) tends to zero at $S \longrightarrow \infty$. Accordingly, we may replace the integral along the straight line $\mathrm{Re}S = \gamma$ in the inverse Laplace transform

$$f(t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} F(S)\exp(St)dS$$

by a contour integral over the left-hand side of the complex plane and apply the residue theorem to the latter.

As a result, the inverse Laplace transform will take the form

$$f(t) = \sum_{j=1}^{+\infty} \operatorname{res}(F(S_n^{(j)})\exp(S_n^{(j)}t)), \qquad (13)$$

where summation is over all roots of the dispersion relations $D_n(S) = 0$ and $S - S_k^{(l)} - S_m^{(g)} = 0$.

Applying formula (13) to coefficient $\xi_n^{(2)}(S)$, we find that

$$\xi_n^{(2)}(t) = \sum_{k,\,m \in \Omega}\sum_{l,\,g,\,j=1}^{+\infty} \frac{\zeta_{kmn}^{lg}(S_n^{(j)}, S_k^{(l)}, S_m^{(g)})}{(S_n^{(j)} - S_k^{(l)} - S_m^{(g)})\partial_{S_n^{(j)}} D_n(S_n^{(j)})}$$

$$\times \exp(S_n^{(j)}t) + \sum_{k,\,m \in \Omega}\sum_{l,\,g=1}^{+\infty} \frac{\zeta_{kmn}^{lg}(S_k^{(l)} + S_m^{(g)}, S_k^{(l)}, S_m^{(g)})}{D_n(S_k^{(l)} + S_m^{(g)})} \qquad (14)$$

$$\times \exp((S_k^{(l)} + S_m^{(g)})t);$$

$$n \ge 2,$$

where $S_n^{(j)}$ is the root of the dispersion relation $D_n(S_n^{(j)})) = 0$.

Coefficients $\xi_0^{(2)}(t)$ and $\xi_1^{(2)}(t)$ will have the form

$$\xi_0^{(2)}(t) = -\frac{1}{r_0}\sum_{m \in \Omega}\sum_{l,\,g=1}^{+\infty} \frac{a_{\xi m}(S_m^{(l)})a_{\xi m}(S_m^{(g)})}{2m+1}$$

$$\times \exp((S_m^{(l)} + S_m^{(g)})t);$$

$$\xi_1^{(2)}(t) = -\frac{9}{r_0}\sum_{m \in \Omega}\sum_{l,\,g=1}^{+\infty} \frac{(m+1)a_{\xi m}(S_m^{(l)})a_{\xi m+1}(S_{m+1}^{(g)})}{(2m+1)(2m+3)}$$

$$\times \exp((S_m^{(l)} + S_{m+1}^{(g)})t).$$

Using the explicit form of coefficients $\xi_n^{(1)}(t)$ and $\xi_n^{(2)}(t)$, as well as expansions (4), (7), and (9), we can write an analytical expression for the generatrix of a nonlinearly vibrating charged drop of a viscous liquid up to second-order terms,

$$r(\vartheta, t) = r_0 + \varepsilon \sum_{n \in \Omega}\sum_{j=1}^{+\infty} a_{\xi n}(S_n^{(j)})\exp(S_n^{(j)}t)P_n(\mu)$$

$$-\frac{\varepsilon^2}{r_0}\sum_{m \in \Omega}\sum_{l,\,g=1}^{+\infty} \frac{a_{\xi m}(S_m^{(l)})a_{\xi m}(S_m^{(g)})}{2m+1}\exp((S_m^{(l)} + S_m^{(g)})t)P_0(\mu)$$

$$-\frac{9\varepsilon^2}{r_0}\sum_{m \in \Omega}\sum_{l,\,g=1}^{+\infty} \frac{(m+1)a_{\xi m}(S_m^{(l)})a_{\xi m+1}(S_{m+1}^{(g)})}{(2m+1)(2m+3)}$$

$$\times \exp((S_m^{(l)} + S_{m+1}^{(g)})t)P_1(\mu) \qquad (15)$$

$$+ \varepsilon^2 \sum_{n=2}^{+\infty}\sum_{k,\,m \in \Omega}\left\{\sum_{l,\,g,\,j=1}^{+\infty} \frac{\zeta_{kmn}^{lg}(S_n^{(j)}, S_k^{(l)}, S_m^{(g)})}{(S_n^{(j)} - S_k^{(l)} - S_m^{(g)})\partial_{S_n^{(j)}} D_n(S_n^{(j)})}\right.$$

$$\times \exp(S_n^{(j)}t) + \sum_{l,\,g=1}^{+\infty} \frac{\zeta_{kmn}^{lg}(S_k^{(l)} + S_m^{(g)}, S_k^{(l)}, S_m^{(g)})}{D_n(S_k^{(l)} + S_m^{(g)})}$$

$$\left. \times \exp((S_k^{(l)} + S_m^{(g)})t) \right\} P_n(\mu).$$

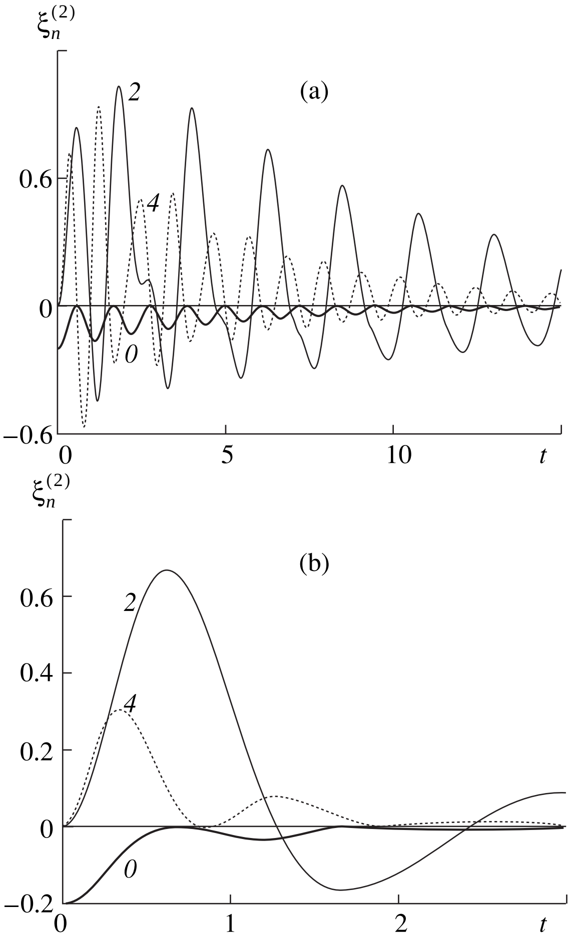


Fig. 1. Dimensional amplitudes of the vibration modes of the liquid viscous drop that are excited in the second order of smallness due to nonlinear interaction vs. dimensionless time at the initial excitation of the fundamental ($n = 2$) mode for $W = 0$ and dimensionless viscosity ν = (a) 0.02 and (b) 0.2. The figure by the curves indicates the mode number. Hereafter, the parameters are made dimensionless at $\rho = r_0 = \sigma = 1$.

Compared with (10), (12), and (14), expression (15) is much more awkward. Here, initially cumbersome coefficients ζ^{lg}_{kmn} (see (12)) are expressed through the integrals of not less cumbersome inhomogeneity functions f^{lg}_{kmn} (i.e., the functions appearing in inhomogeneous equation (10)). Thus, direct use of expression (15) seems problematic. Even with simple (single-mode) initial deformations of the drop, the calculation procedure is a big bugaboo even if the modes excited are not too high (i.e., even though the spectrum of modes excited via nonlinear interaction is relatively narrow). Therefore, a solution to (15) is illustrated (Fig. 1) for the nonlinear vibration of a viscous liquid drop when its initial deformation is specified by the lowest (fundamental) mode ($n = 2$).

(5) An important issue related to nonlinear vibrations of a viscous liquid drop is finding the viscosity dependence of the spectrum of modes excited through nonlinear interaction. Note, first of all, that the spectrum of these modes depends on inhomogeneity function f^{lg}_{kmn} on the right of (10), through which coefficients ζ^{lg}_{kmn} (defining the second-order corrections in (15)) are expressed in the integral form.

Performing calculations by (15), we revealed that the inhomogeneity function in differential equation (10) may take a simpler form, since coefficients f^{1}_{kmn}, f^{3}_{kmn}, f^{5}_{kmn}, and f^{7}_{kmn} have the properties

$$f^{1}_{kkn} = 0, \quad f^{1}_{kmn} = -f^{1}_{mkn}, \quad f^{3}_{kmn} = 0, \quad f^{5}_{kmn} = 0,$$
$$f^{7}_{kkn} = 0, \quad f^{7}_{kmn} = -f^{7}_{mkn},$$

at least if any set of modes with numbers from 2 to 20 is initially excited (i.e., for any k and $m \in \Omega$ and $n \geq 2$ when $\Omega \subseteq \Omega_0$ and $\Omega_0 = \{k|\, 2 \leq k \leq 20, k \in N\}$). It should be noted that this mode interval is of the greatest interest, since higher modes decay quickly and have no essential effect in physically significant applications. With regard to the above relationships, function $f^{lg}_{kmn}(r, S^{(l)}_k, S^{(g)}_m)$ on the right of (10) can be represented in the more compact form

$$\begin{aligned} f^{lg}_{kmn}(r, S^{(l)}_k, S^{(g)}_m) \;=\; & f^{4}_{kmn}(\chi^{(g)}_m)^3 a_k(S^{(l)}_k) b_m(S^{(g)}_m)\frac{r^{k-2}}{r_0^{k-1}} \\ & \times \frac{j_{m+1}(\chi^{(g)}_m r)}{j_m(\chi^{(g)}_m r_0)} + f^{2}_{kmn}(\chi^{(g)}_m)^2 b_m(S^{(g)}_m)\left(a_k(S^{(l)}_k)\frac{r^{k-3}}{r_0^{k-1}} \right. \\ & \left. + b_k(S^{(l)}_k)\frac{1}{r^3}\frac{j_k(\chi^{(l)}_k r)}{j_k(\chi^{(l)}_k r_0)}\right)\frac{j_m(\chi^{(g)}_m r)}{j_m(\chi^{(g)}_m r_0)} + (f^{4}_{kmn}(\chi^{(g)}_m)^3 \\ & + f^{6}_{kmn}(\chi^{(l)}_k)^2\chi^{(g)}_m) b_k(S^{(l)}_k) b_m(S^{(g)}_m)\frac{1}{r^2}\frac{j_k(\chi^{(l)}_k r)}{j_k(\chi^{(l)}_k r_0)}\frac{j_{m+1}(\chi^{(g)}_m r)}{j_m(\chi^{(g)}_m r_0)}. \end{aligned}$$

If at least one mode is initially excited, the zeroth-order mode will be excited in the second order of smallness because of the constancy of the volume of the drop. The first mode in the second order of smallness will be excited only if two neighboring modes (with their numbers differing by unity) are excited in the first order of smallness [5, 6]. The higher mode amplitudes ($n \geq 2$), which are given by (15), are proportional to coefficients K_{kmn}, α_{kmn}, Γ_{kmn}, Γ_{mkn}, Λ_{kmn}, and Λ_{mkn}. The first four of them are proportional to Clebsch–Gordan coefficients

$$C^{n0}_{k0m0}, C^{k0}_{n0m0}, C^{m0}_{n0k0}; \quad k, m \in \Omega; \quad n \geq 2, \tag{16}$$

and the remaining two are representable as a sum of two terms: one is a sum of type

$$\sum_{j=1}^{[m/2]} \alpha_{n,k,m-2j}, \quad \sum_{j=1}^{[k/2]} \alpha_{n,m,k-2j}; \quad k, m \in \Omega; \quad n \geq 2, \tag{17}$$

and the other is proportional to coefficients C^{k0}_{n0m0} and C^{m0}_{n0k0}. As for coefficients (16), they are other than zero only if $k + m + n$ is an even number and the triangle inequality $|k - m| \leq n \leq k + m$ is fulfilled. Coefficients

$$\begin{aligned} &\alpha_{n,k,m-2j}, \quad j = 1, 2, \ldots, [m/2]; \\ &\alpha_{n,m,k-2j}, \quad j = 1, 2, \ldots, [k/2]; \quad k, m \in \Omega \end{aligned} \tag{18}$$

are other than zero if $\min\{|k - m + 2j|\} \leq n \leq k + m - 2$ ($j = 1, 2, \ldots, m/2$) and $\min\{|m - k + 2j|\} \leq n \leq m + k - 2$ ($j = 1, 2, \ldots, k/2$). For example, if one mode with number n_1 (i.e., $\Omega = \{n_i\}$) is initially excited, coefficients (16) will be other than zero for even n satisfying the inequality $2 \leq n \leq 2n_1$ and coefficients (18), for n such that $2 \leq n \leq 2n_1 - 2$. Hence, the zero mode and even modes with numbers from the range $2 \leq n \leq 2n_1$ will be excited in the second order of smallness in this case. Now let two modes with numbers n_1 and n_2 ($\Omega = \{n_1, n_2\}$) be initially excited. Then, if both n_1 and n_2 are even or odd, only those coefficients given by (16) and (18) will be nonzero for which n is even and satisfies the conditions $2 \leq n \leq \max\{2n_1, 2n_2\}$ and $2 \leq n \leq \max\{2n_1, 2n_2\} - 2$. If one of numbers n_1 and n_2 is even and the other odd, those coefficients given by (16) will be nonzero for which n either is even and satisfies the inequality $2 \leq n \leq \max\{2n_1, 2n_2\}$ or is odd and such that $|n_1 - n_2| \leq n \leq n_1 + n_2$. Of coefficients (18) those will be other than zero for which n is either even and meets the inequality $1 \leq n \leq \max\{2n_1, 2n_2\} - 2$ or is odd and such that $3 \leq n \leq n_1 + n_2 - 2$. Note that the last inequality for odd n is formulated for expressions (18). Actually, however, for modes with odd numbers such that $3 \leq n \leq n_1 + n_2 - 2$ be present in the second-order spectrum, it is necessary that coefficients f^{lg}_{kmn} in (10) be other than zero. Looking through values of n_1 and n_2 of which one is even and the other odd in the interval from 2 to 20, we get the conclusion that, for $3 \leq n \leq |n_1 - n_2|$, coefficients (17) are always zero, since sums (17) always have terms that cancel out.

Thus, for a viscous liquid drop, the spectrum of modes excited in the second order of smallness through the nonlinear interaction coincides with that for a drop of an ideal liquid [7–9].

(6) From expression (15), it follows that the solution to the problem of capillary vibration of a viscous incompressible liquid is of a resonant character: the denominators of some of the terms in the analytical expression for the generatrix of the drop become small under definite conditions. In particular, for nonlinearly interacting modes, a combined three-mode resonance may take place at $\mathrm{Re}S_n^{(j)} \ll \mathrm{Im}S_n^{(j)}$, when the imaginary part of the difference between the dimensionless (at $\rho = 1$, $r_0 = 1$, $\sigma = 1$) frequencies in the denominator vanishes, $\mathrm{Im}(S_n^{(j)} - S_k^{(l)} - S_m^{(g)}) = 0$, where $S_n^{(j)}$ is the root of the dispersion relation $D_n(S_n^{(j)}) = 0$. It should be remembered that, for an ideal liquid, the resonance condition sets in when the difference between the (real) frequencies of interacting modes vanishes.

The dispersion relation $D_n(S_n^{(j)}) = 0$ has an infinite number of roots, among which two are complex conjugate at a subcritical (in terms of stability against self-charge) value of W: $S_n^{(1)} = -\delta_n + i\varpi_n$ and $S_n^{(2)} = -\delta_n - \varpi_n$, where $\delta_n = -\mathrm{Re}(S_n^{(1)}) = -\mathrm{Re}(S_n^{(2)})$ is the damping decrement and $\varpi\omega_n = \mathrm{Im}(S_n^{(1)}) = -\mathrm{Im}(S_n^{(2)})$ is the vibration frequency. The rest of the roots of the dispersion relation $D_n(S_n^{(j)}) = 0$ are real and negative. At a supercritical value of W, all the roots are real and at least one of them is larger than zero. Consequently, at subcritical W, resonance may take place if $\varpi_n = \varpi_k + \varpi_m$. Under resonance, the denominators of the terms in series (15) will contain expressions $S_n^{(1)} - S_k^{(1)} - S_m^{(1)} = S_n^{(2)} - S_k^{(2)} - S_m^{(2)} = -\delta_n + \delta_k + \delta_m$, whose numerical values depend on the viscosity and tend toward zero at $\nu \longrightarrow 0$.

Let us take a closer look at two degenerate nonlinear resonances involving the fourth and sixth and the fifth and eighth modes. Degenerate resonances occur if only one mode, e.g., with number k is initially excited. The resonance condition in this case is

$$\varpi_n = 2\varpi_k. \tag{19}$$

From (19), one can find a relationship between the resonance value of W, W_r, and viscosity ν (Fig. 2). It is seen that resonance value W_r of the Rayleigh parameter grows with viscosity at least for degenerate resonances.

Minimal indices of summation in infinite series (14) at which it can be fairly accurately fitted by a finite number of its initial terms at $n = k = m = 2$ and different values of dimensionless (at $\rho = r_0 = \sigma = 1$) viscosity ν and parameter W

ν	$W = 0$	$W = 2$	$W = 3$	$W = 3.5$
0.01	22	20	15	10
0.02	16	12	9	6
0.03	10	8	7	5
0.05	8	6	5	4
0.07	6	5	4	3
0.1	4	4	3	3
0.2	3	3	3	3
0.5	3	3	3	2
1.5	2	2	2	2
2.0	2	1	1	1

(i) For a low-viscosity liquid, the above statement is strengthened by approximate calculation under the assumption that [4]

$$j_n(\chi) = \frac{\exp(\chi)}{2\chi}\left(1 - \frac{n(n+1)}{2\chi} + \frac{n(n^2-1)(n+2)}{8\chi^2} + O\left(\frac{1}{\chi^3}\right)\right); \quad \chi \longrightarrow \infty.$$

With this assumption, we can find an approximate expression for the first root of the dispersion relation $D_n(S_n^{(j)}) = 0$ in the form $S_n^{(1)} = -\delta_n + i\varpi_n$, where damping decrement δ_n and frequency ϖ_n are given by

$$\delta_n = (n-1)(2n+1)\frac{\nu}{r_0^2} - \sqrt{2}(n-1)^2(n+1)\frac{\nu^{3/2}}{r_0^3\sqrt{\omega_n}} + o(\nu^2); \tag{20}$$

$$\varpi_n = \omega_n - \sqrt{2}(n-1)^2(n+1)\frac{\nu^{3/2}}{r_0^3\sqrt{\omega_n}} - (n-1)^2(4n+5)\frac{\nu^2}{2r_0^4\omega_n} + o(\nu^2). \tag{21}$$

Substituting (21) into (19), one easily finds the viscosity dependence of W_r,

$$W_\mathrm{r} = W_\mathrm{r}^{(0)} + \frac{8\sqrt{2}(k-1)kn(k-n)(n-1)(2(k-1)^2(k+1)\sqrt{\omega_n^{(0)}} - (n-1)^2(n+1)\sqrt{\omega_k^{(0)}})}{r_0^3\sqrt{\omega_n^{(0)}}\sqrt{\omega_k^{(0)}}(4k(k-1) - n(n-1))(8(k-1)k\omega_k^{(0)} - (n-1)n\omega_n^{(0)})}\nu^{3/2} - \frac{4k(k(4k-3)-6) - n(n(4n-3)-6) + 15}{4(k-1)k - (n-1)n}\frac{\rho\nu^2}{r_0\sigma}. \tag{22}$$

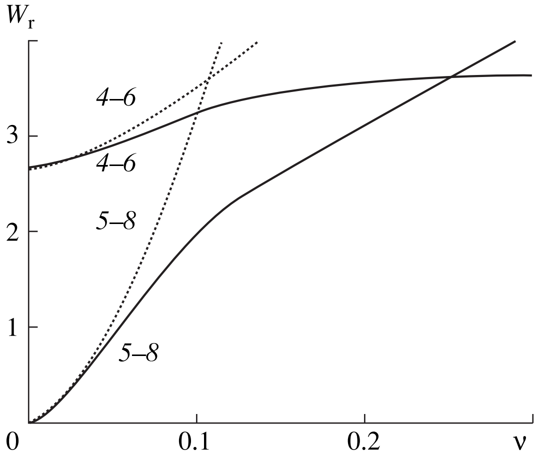


Fig. 2. Resonance value of parameter W vs. dimensionless viscosity ν for degenerate resonances between (*4–6*) the fourth and sixth modes and (*5–8*) fifth and eighth modes. The continuous curve, exact solution; the dotted curve, calculation by (15).

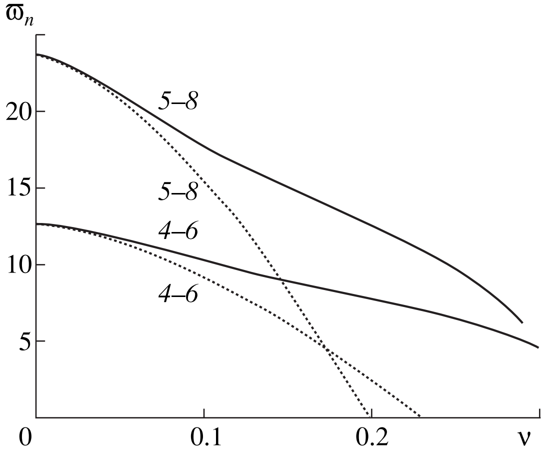


Fig. 3. Resonance frequency ϖ_n vs. dimensional viscosity ν. The continuous curve, exact solution; the dotted curve, calculation by (17). For the figures by the curves, see Fig. 1.

Here, $W_r^{(0)}$ and $\omega_n^{(0)}$ are the resonant value of parameter W and the capillary vibration frequency for an ideal liquid,

$$W_r^{(0)} = \frac{4k(k-1)(k+2)-n(n-1)(n+2)}{4k(k-1)-n(n-1)};$$

$$\omega_n^{(0)} = \sqrt{\sigma n(n-1)(n+2-W_r^{(0)})/(\rho r_0^3)}.$$

Substituting (22) into (20) yields approximate expressions for the resonant values of the damping decrement and vibration frequency,

$$\delta_n = (n-1)(2n+1)\frac{\nu}{r_0^2} - \sqrt{2}(n-1)^2(n+1)\frac{\nu^{3/2}}{r_0^3\sqrt{\omega_n^{(0)}}};$$

$$\varpi_n = \omega_n^{(0)} - \frac{2\sqrt{2}(k-1)(n-1)(n(k^2-1)(\omega_n^{(0)})^{3/2} - 4k(n^2-1)(\omega_k^{(0)})^{3/2})}{(n(n-1)\omega_n^{(0)} - 8k(k-1)\omega_k^{(0)}\sqrt{\omega_k^{(0)}}\sqrt{\omega_n^{(0)}})}\frac{\nu^{3/2}}{r_0^3}$$
$$-\left(\frac{(n-1)^2(4n+5)}{2\omega_n^{(0)}r_0^4} + \frac{4k(k(4k-3)-6)-n(n(4n-3)-6)+15}{8(k-1)k(k-n)}\frac{\rho\omega_n^{(0)}}{\sigma r_0}\right)\nu^2. \qquad (23)$$

Expressions (20)–(23) adequately describe the viscosity dependences of the physical parameters only when the viscosity is low, as follows from Figs. 2 and 3. Under resonance, the differences between the complex frequencies, $S_n^{(1)} - 2S_k^{(1)} = S_n^{(2)} - 2S_k^{(2)} = -\delta_n + 2\delta_k$ remain finite at $\nu \neq 0$, indicating the finiteness of the resonating mode amplitudes (Figs. 4 and 5).

Viscosity affects not only the resonance frequencies and the resonance values of the Rayleigh parameter but also the convergence of series (14). As the viscosity grows, the minimal number of the roots of the dispersion relation that provide a good convergence of series (14) (i.e., the minimal number of terms that can approximate infinite series (14) to such a degree that the difference between the exact and approximate solutions is within the thickness of the lines in Figs. 4 and 5) decreases drastically (see table). This necessitates the search for a more compact expression for $\xi_n^{(2)}(t)$ at moderate and high viscosity values, i.e., under the conditions where one or two (according to the value of ν) roots of the dispersion relation (one or two terms of series (14)) will suffice (the case of two terms is exemplified in [1]).

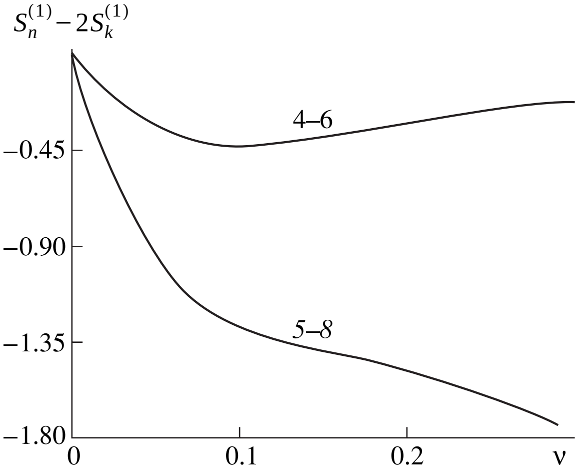


Fig. 4. Difference $S_n^{(1)} - 2S_k^{(1)}$ between the dimensionless frequencies vs. dimensional viscosity ν. For the figures by the curves, see Fig. 1.

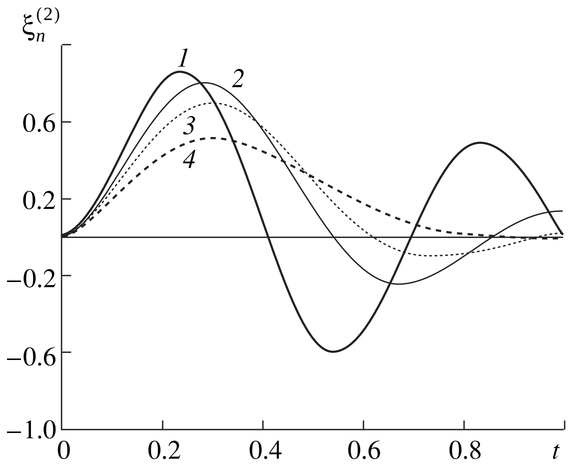


Fig. 5. Dimensionless coefficient $\xi_n^{(2)}$ vs. dimensionless time t calculated by (14) for resonating fourth and sixth modes at ν = (*1*) 0.05, (*2*) 0.1, (*3*) 0.15, and (*4*) 0.25.

CONCLUSIONS

Considering a real (viscous), rather than ideal, liquid drop in the problem of nonlinear vibration does not extend the spectrum of modes specifying the shape of its surface (compared with the ideal case) but increases the resonance value of the self-charge and limits the mode amplitudes in exact resonance.

ACKNOWLEDGMENTS

This work was supported by a grant of the President of the Russian Federation (grant no. MK-2946-2004-1) and the Russian Foundation for Basic Research (grant no. 03-01-00760).

REFERENCES

1. A. N. Zharov and A. I. Grigor’ev, Zh. Tekh. Fiz. **75** (1), 22 (2005) [Tech. Phys. **50**, 19 (2005)].
2. S. O. Shiryaeva, D. F. Belonozhko, V. B. Svetovoi, and A. I. Grigor’ev, Preprint No. 31, IMI RAN (Institute of Microelectronics, Yaroslavl, 2001).
3. A.-H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973; Mir, Moscow, 1976).
4. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).
5. S. O. Shiryaeva, Zh. Tekh. Fiz. **72** (4), 15 (2002) [Tech. Phys. **47**, 389 (2002)].
6. 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)].
7. J. A. Tsamopolous and R. A. Brown, J. Fluid Mech. **127**, 519 (1983).
8. S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 173 (2001).
9. A. N. Zharov, S. O. Shiryaeva, and A. I. Grigor’ev, Zh. Tekh. Fiz. **73** (12), 9 (2003) [Tech. Phys. **48**, 1511 (2003)].

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 846–855. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 29–39.
Original Russian Text Copyright © 2005 by Lokhman, Ogurok, Ryabov.

Formation and Measurement of Cluster Beams at Gasdynamic Cooling of CF_2HCl Molecules

V. N. Lokhman, D. D. Ogurok, and E. A. Ryabov

Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow Oblast, Russia

e-mail: ryabov@isan.troitsk.ru

Received May 17, 2004

Abstract—The clustering of CF_2HCl (Freon-22 refrigerant) molecules is found to take place when these molecules are dynamically cooled in a pulsed supersonic beam. A method of CF_2HCl cluster beam diagnostics is developed that combines UV multiphoton ionization, time-of-flight mass spectroscopy, and cluster IR photodissociation. The velocity of directed motion of $(CH_2HCl)_n$ clusters, as well as the longitudinal and transverse velocities of their thermal motion, are measured at different stagnation pressures P_0. The cluster mean size and the degree of clustering in the flow are estimated depending on supersonic flow conditions. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

It has been proved thus far that laser separation of ^{13}C and ^{12}C carbon isotopes, which is based on IR multiphoton dissociation of Freon-22 molecules under the action of CO_2 laser pulsed radiation, is feasible [1]. Research is now underway that is aimed at improving the yield and selectivity of laser isotope separation. One way to improve the selectivity is narrowing the IR absorption spectrum of the molecules via gasdynamic cooling in pulsed supersonic beams [2–7]. However, profound cooling may cause clustering in the beam [8, 9] and change its parameters to the point where the method of laser isotope separation loses isotope selectivity. In light of this, the development of techniques measuring Freon-22 parameters in supersonic molecular beams (specifically, under clustering conditions) seems topical. One such technique using a time-of-flight ionization mass spectrometer is presented in this work. Preliminary studies of CF_2HCl molecule clustering are reported elsewhere [10].

Gasdynamic cooling and clustering taking place in a supersonic gas flow into a vacuum have received much attention (see, e.g., [8, 9, 11–13]). It has been shown that these processes depend on the thermodynamic properties of the gas (or gas mixture), stagnation parameters (T_0, P_0), and nozzle geometry (diameter) and design (the presence or absence of an extension). Clustering occurs if (i) the flow is rapidly cooled passing into the supersaturated nonequilibrium state; (ii) condensation centers are available: in the homogeneous gas, these are ultrasmall ($n < 5$, where n is the number of molecules) clusters produced by three-body collisions; and (iii) there exists a collision area where the relatively slow condensation process is observed, which terminates when the collisionless flow sets in.

The main goal of our study is to find CF_2HCl molecule condensation conditions and measure the basic cluster flow (beam) parameters (degree of condensation q; cluster mean size N; velocity U of directed motion; and longitudinal, $V_{\|}$, and transverse, $V_{\perp}$, components of the thermal velocity of clusters in the beam) depending on supersonic flow conditions.

1. MEASURING SEUUP AND TECHNIQUE

The measuring setup was described in detail elsewhere [14]. All experiments were carried out with pure (without a carrier gas) Freon-22 flowing out from a supersonic nozzle without a conic extension (at temperature $T_0 = 298$ K). The measuring geometry is shown in Fig. 1a. In the source chamber, skimmer Sk (Beam Dynamics Co., $D_S = 0.66$ mm) placed at distance $\Delta Z_{NS} = 38.5$ mm from pulsed electromagnetic nozzle *PN* (General Valve Co., $d = 0.8$ mm, $\Delta t = 200$ μs) cuts a molecular beam from a supersonic flow (jet) produced by the nozzle. The molecular beam thus formed enters the chamber of a mass spectrometer, where it meets the mutually orthogonal axes of the mass spectrometer (the Y axis) and focused ($f = 12$ cm) pulsed (a pulse width of 7 ns) ionizing UV ($\lambda \approx 232$ nm) laser radiation at distance $\Delta Z_{SD} = 96.5$ mm from the entrance to the skimmer. The vibrational excitation of the particles is accomplished with 150-ns-wide pulses from a CO_2 laser. Varying the delay time between nozzle switching and UV pulse application, one can record time-of-flight (TOF) spectrum $S(t, y = 0)$ of the beam, i.e., the time dependence of the beam molecule concentration as the molecules pass through the detection area. (Fig. 1b). With the delay fixed, particle distribution $S(y)$

over the cross section of the beam can be taken by moving the detection area along the *Y* axis. Taken together, distributions $S(t, y = 0)$ and $S(y)$ allowed us to derive the space–time characteristics of the beam versus the laser radiation parameters and beam formation parameters (T_0, P_0) for signals with different mass numbers, $S(t, y; T_0, P_0; E_{UV}, \lambda_{UV}; \Phi_{IR}, \nu_{IR}; M/e)$.

The onset of developed condensation shows up usually as a sharp increase in the beam intensity versus stagnation pressure curve [15, 16]. In addition, the presence of clusters in our case was detected from the breakdown of the cluster component due to IR radiation (see below). In the course of experiments, multiphoton fragmentation and ionization of the particles was observed when the UV radiation intensity was sufficiently high. One of the basic photoionization products in this case are ions with peaks at $M/e = 31$ (CF^+) and $M/e = 51$ (CF_2H^+). Special measurements showed that both monomers and clusters contribute to signal $S(31)$, while signal $S(51)$ (at E_{UV} = 100–270 μJ, Φ_{UV} = 10–27 J/cm^2), $\lambda = 232$ nm) is due to monomers alone. Accordingly, one can separately trace the behavior of monomers and clusters in the beam.

Velocity *U* of directed motion of clusters was determined from the difference between time t_D the front of a step disturbance (“mark”) arrives at the detection zone with coordinate Z_D and time t_{IR} of generation of this disturbance at the entrance to the skimmer with coordinate Z_S, $U = \Delta Z_{SD}/(t_D - t_{IR})$ (Fig. 1b). The mark resulted from IR photodissociation of clusters near the skimmer when the radiation of the CO_2 laser resonantly acted on vibrations ν_3 and ν_8 of CF_2HCl molecules (ν_{las} = 1037.4 cm^{-1}). The energy density of the IR radiation was $\Phi_{IR} \approx 0.4$ J/cm^2, which is much lower than the IR multiphoton dissociation threshold for CF_2HCl molecules ($\Phi_{th} \approx 3$ J/cm^2) [7]. The transverse, $V_{\parallel}$, and longitudinal, $V_{\perp}$, components of the thermal velocity of clusters in the beam were determined by comparing the experimental and analytical data: $V_{\parallel}$ was estimated from the smearing of the mark front along the *Z* axis (Fig. 1b) in spectrum $S(t, y = 0)$; $V_{\perp}$, from the evolution of corresponding distributions $S(y)$ over the cross section of the beam (Fig. 1a).

Fraction *q* of condensed molecules in the jet and mean cluster size *N* were found from the set of the above dependences taken at different stagnation pressures.

We also measured efficiency ΔQ of the pulsed nozzle as a function of stagnation pressure P_0 (see Fig. 2). With regard to the gas flow rate, the estimated effective cross-sectional area of the nozzle turns out to be 30 times as small as its clear area. When the valve opens, the gas is likely to flow out through a narrow annular slit between valve *1* and Teflon cock plug *2* into a short cylinder (Fig. 2) and gasdynamic cooling starts in this cylinder, continuing as the jet further expands in the source chamber. Such flow conditions seem to be the reason for the nonlinear dependence of the nozzle

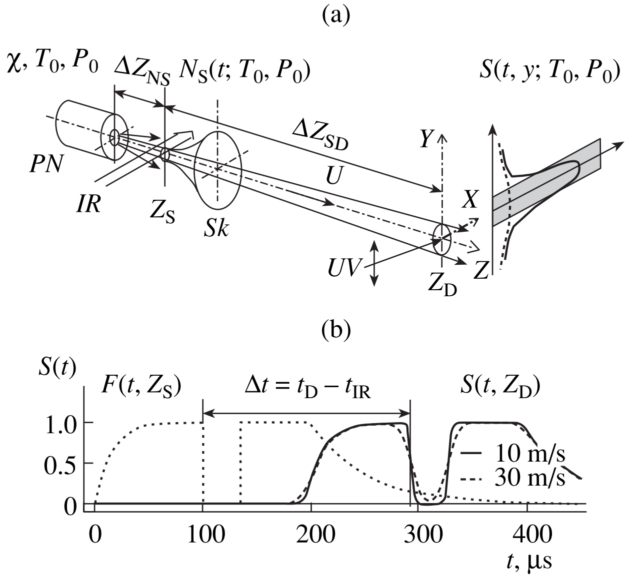


Fig. 1. (a) Geometry of measurement: *PN*, pulsed nozzle; *Sk*, skimmer; *IR* and *UV*, beams of *IR* and *UV* pulsed laser radiation; the *Y* axis is aligned with the axis of the TOF mass spectrometer; and $S(t, y; T_0, P_0)$, transverse distribution of the signal (in the *Y* direction). (b) Simulation of the TOF spectrum: $F(t, Z_S)$, the source function with a photodissociative mark at time instant t_{IR} applied to the entrance to the skimmer; $S(t, Z_D)$, calculated TOF signal near the detector at $V_{\parallel}$ = 10 and 30 m/s and U = 500 m/s.

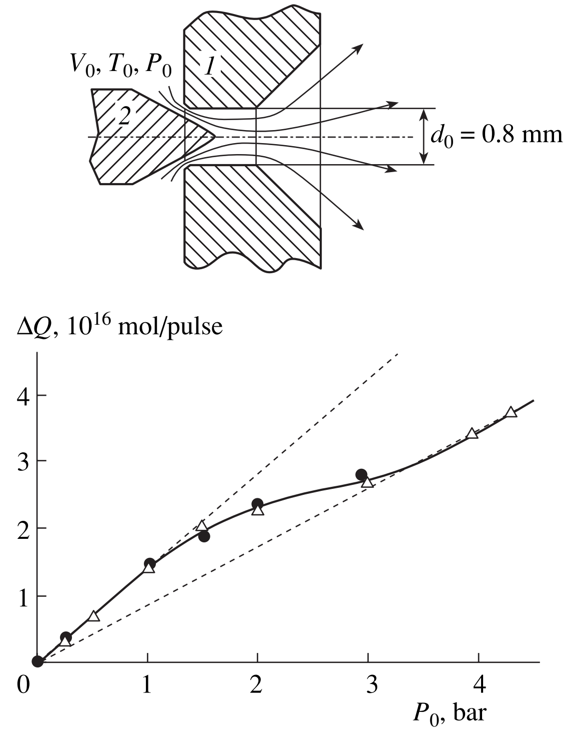


Fig. 2. Mass flow rate per pulse vs. the pressure. At the top, part of the pulsed valve is shown: (*1*) body of the valve and (*2*) Teflon cock plug.

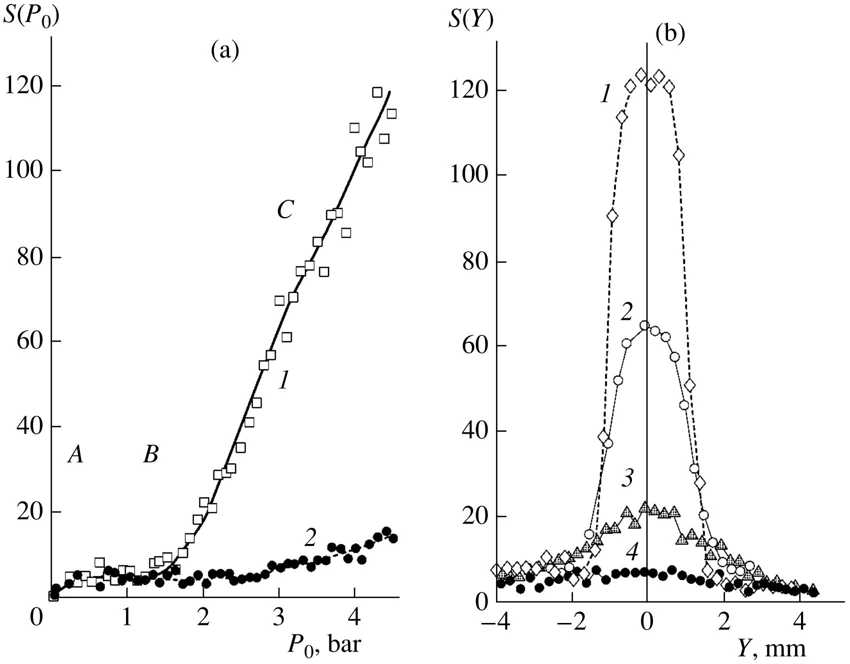


Fig. 3. (a) Photoionic signals $S(P_0,M/e)$ for mass numbers M/e = (*1*) 31 (CF^+) and (*2*) 51 (CF_2H^+) vs. stagnation pressure. (b) Transverse distributions $S(y)$ for (*1–3*) P_0 = 4.5, 3.0, and 2.0 bar, respectively, and M/e = 31 and (*4*) P_0 = 2.0 bar and M/e = 51. The apparatus functions are disregarded.

efficiency on the stagnant pressure ($\Delta Q(P_0)$ in Fig. 2). Since no appreciable variation of the pulse duration was observed in the CF_2HCl pressure range used (0.1–4.5 bar), so that a pressure shock in the chamber of the mass spectrometer is proportional to ΔQ, one may assume that the total concentration of molecules in the jet is proportional (at least at the entrance to the skimmer, N_S) to the gas mass rate per pulse, $N_S \propto \Delta Q(P_0)$, rather than directly to P_0. This circumstance was taken into account in analyzing the results of measurement.

2. RESULTS OF MEASUREMENT

The experimental dependences of the photoionic signal on stagnation pressure P_0 for $S(P_0, M/e = 31)$ and $S(P_0, M/e = 51)$, as well as transverse distributions $S(y)$ for several stagnation pressures, are shown in Figs. 3a and 3b.

The curve $S(P_0, M/e = 31)$ (curve *1* in Fig. 3a) is typical of the condensation process, specifically, of condensation of molecular gases (CO_2, SF_6, CF_2Cl_2) in expanding jets [15, 16]. As the pressure exceeds a certain value (marked by arrow *B* in Fig. 3a), the intensity of the central part of the beam increases greatly, reflecting the formation of stable larger-than-critical clusters (n = 5–10 or higher). At this stage (developed condensation), cluster nuclei grow largely by molecule condensation via two-body collisions when the jet is supersaturated. At the precondensation stage (lower pressure portion *A*–*B*), ultrasmall ($n < 5$) clusters nucleate via three-body collisions, which favor removal of the latent heat of vaporization and the vibrational excitation energy, enhancing the rate of vibrational (*V*–*T*) relaxation [17].

Comparing the pressure dependences of the photoionic signal at different masses (M/e = 31 and 51, curves *1* and *2* in Fig. 3a), one can see that general curve *1* (including the cluster component) and partial (monomer) curve *2* behave in a different fashion. In addition, Fig. 3b shows that, at pressures above P_B, the sharp increase in the intensity is accompanied by narrowing of the beam's cross-sectional area: its width approaches that defined by the geometry of the experiment (i.e., by the diameter and position of the skimmer). As will be shown later, such behavior is due mainly to the cluster component of the signal. Signal $S(51)$, corresponding to the monomer component, has a much lower intensity, as well as exhibits a less pronounced pressure dependence and a wider distribution throughout the pressure range used (in Fig. 3b, signal $S(51)$ is shown only for P_0 = 2 bar).

As was noted, the contribution of the monomer component to $S(M/e = 31)$ can be taken into account, making it possible to separate out the corresponding dependences for the cluster component. Figure 4 plots the cluster component of the signal against the mass flow rate per pulse. This curve was constructed based on Fig. 2. The rapid rise of the signal in portion *B*–*C* (see below) is due both to a rise in the total concentration of clusterized molecules in the beam and to a decrease in the transverse diffusion of the beam. The latter fact leads to a fast growth of the particle concen-

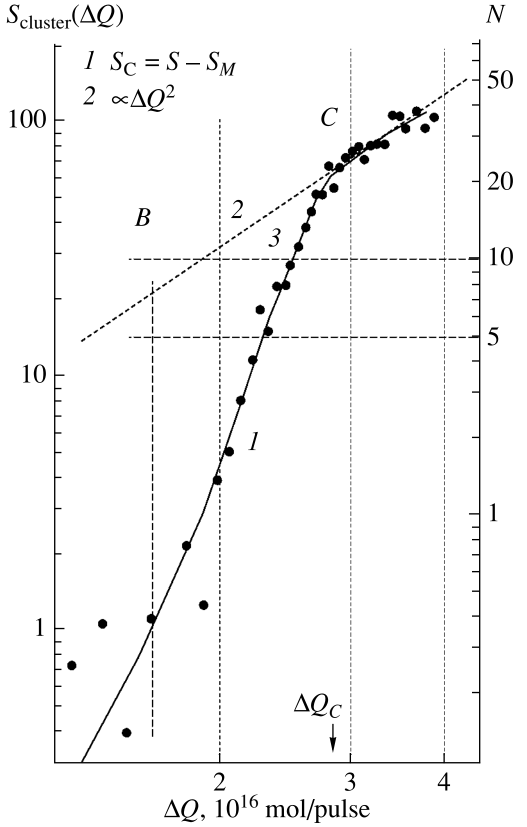


Fig. 4. (*1*) Cluster component of the signal vs. mass flow rate per pulse; (*2*) total concentration of molecules entering into the clusters near the skimmer vs. mass flow rate per pulse, $qN_S \propto \Delta Q^2$; and (*3*) estimated mean cluster size (the vertical axis on the right) vs. mass flow rate per pulse, $N(\Delta Q) = N_B(\Delta Q/\Delta Q_B)^2$ (the corresponding curve coincides with dashed curve *2*).

tration in the central part of the beam, because the transverse component of the thermal velocity is to a great extent suppressed. As the pressure rises further, above point *C*, the diffusion factor loses significance and the dependence $S_C(\Delta Q)$ characterizes the run of the molecule concentration in the cluster near the skimmer, qN_S, as a function of the mass flow rate per pulse. Here, N_S is the molecule concentration in the jet near the skimmer and q is the fraction of molecules condensed. It is seen that $qN_S \propto S_C(\Delta Q) \propto (\Delta Q)^2$ in this region; hence, the degree of clustering is $q \propto \Delta Q$.

The velocity characteristics of the longitudinal motion of the cluster beam (U and $V_\parallel$) were found from modulated TOF spectra $S(t, y = 0)$. Fragments of the TOF spectra for the cluster component of the signal near the front of the step mark that were taken at several pressures are shown in Fig. 5, which also shows the corresponding model spectra for comparison. Both spectra are seen to be in good agreement. Below, the problem of determining the beam parameters will be considered in greater detail.

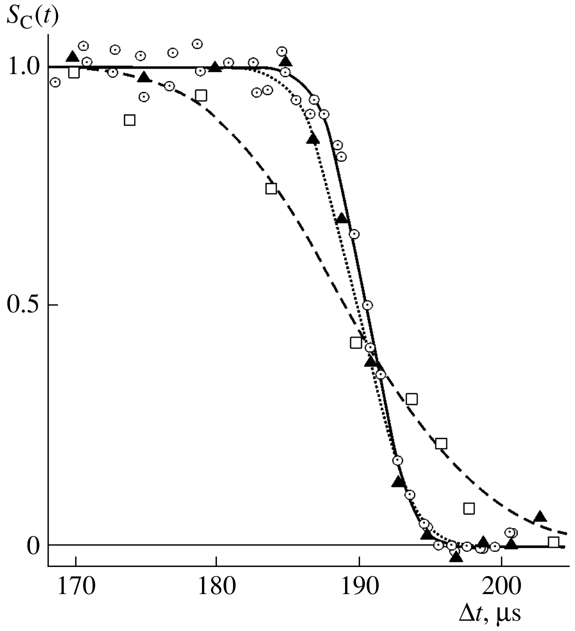


Fig. 5. Normalized experimental and calculated TOF spectra for the cluster component of the beam near the front of the step mark for P_0 = (□) 2.0, (▲) 3.25, and (⊙) 3.5 bar. Δt is the time delay between the probing UV pulse and the IR photodissociating pulse. The values of the directed motion velocity determined from the time of arrival of the front (at its half-height) are U = 509.5, 507.4, and 505.6 m/s, respectively. The calculated spectra were taken for the longitudinal component of the thermal velocity of clusters $V_\parallel$ = 30.4, 9.4, and 7.3 m/s, respectively.

3. SIMULATION AND ANALYSIS OF EXPERIMENTAL DATA

When processing and analyzing the experimental data, we used the following model. It is assumed that the total concentration of molecules at the entrance to the skimmer, N_S, is defined by the contribution of monomers and molecules entering into the clusters: $N_S(\Delta Q(P_0)) = (1 - q)N_S + qN_S$. It is also assumed that the molecules of the cluster component are distributed among clusters of different size n with certain normalized distribution $f(n)$: $\sum_n f(n) = 1$. Then, the concentration of clusters of given size n is given by

$$C_n = f(n)qN_S/\sum_n nf(n) = f(n)qN_S/N,$$

where N is the cluster mean size and nC_n is the concentration of molecules in these clusters, $\sum_n nC_n = qN_S$.

In our experiments, the measured amplitude of signal $S(M/e = 31)$ is defined by the total concentration of molecules in the detection volume,

$$S(t, y;\, T_0, P_0) = S_1 + S_C$$

$$= \zeta_1\varphi_1^2 B_1(t, y;\, U_1, V_1)(1-q)N_S$$

$$+\sum_n \zeta_n \varphi_n^2 B_n(t, y; U_n, V_n) n C_n. \tag{1}$$

Here, ζ_1 and ζ_n are the parameters taking into account the ionization efficiency and apparatus functions for monomers and n-molecule clusters, respectively; $\varphi^2 = N_D/N_S = (D_S/2R_0)^2$ (where D_S is the diameter of the entrance opening of the skimmer) is a geometrical factor allowing for a decrease in the concentration because of beam divergence; R_0 is the geometrical radius of the beam near the detector; and $B_n(t, y; U_n, V_n)$ are the respective normalized functions that describe the space-dependent concentration of particles of different size n at their collisionless motion in the post-skimmer space according to their velocity distributions $f_n(v)$. This model uses the elliptical drift Maxwellian distribution [6]

$$f(v) \propto \left(\frac{1}{\sqrt{\pi} V_\parallel}\right) \times \exp\left(\frac{-(v_\parallel - U)^2}{V_\parallel^2}\right)\left(\frac{1}{\pi V_\perp^2}\right)\exp\left(-\frac{v_\perp^2}{V_\perp^2}\right), \tag{2}$$

where U is the velocity of directed motion of the beam and $V_\parallel$ and $V_\perp$ are the corresponding components of the most probable thermal velocity of beam particles.

To correctly simulate the signal due to the cluster component, one must know a specific form of cluster size distribution $f(n)$ and of velocity distribution $V(n)$ versus particle size in an ensemble of clusters (with ΔQ fixed). The distribution $V = V(n)$ implies that, by the time the jet switches to collisionless flow, the energy distribution over particles of different size is nonequilibrium. Here, one can separate two limiting cases.

(i) The situation is essentially nonequilibrium. Then, as was demonstrated for small water clusters ($5 < n < 40$), the velocity of random motion under the same flow conditions for clusters of different size (within a given ensemble) can be characterized by single parameter $V_n = \sqrt{(2nkT^*)/(nm)} = \sqrt{2kT^*/m} = V^*(\Delta Q)$, where m is the molecular mass [18]. Hence, it follows that the temperature depends on the cluster size, $T_n = nT^*$. Basically, the values of T^* may differ for $V_\parallel$ and $V_\perp$.

(ii) The situation is totally equilibrium. In this case, the particles of different size have the same temperature, $T_n = T(\Delta Q)$, while their velocities are size-dependent, $V(n) \propto 1/\sqrt{n}$.

As follows from [18], where experimental ultrasmall ($n < 5$) and small ($5 < n < 40$) water cluster distributions were obtained, the general small particle size distributions may be fairly accurately approximated by a sum of two exponentials (for either subsystem), each with its own weights and mean size parameter (N). In our experiments, the contribution of ultrasmall clusters to the signals measured is insignificant. Supposedly, this is because the number of molecules in such clusters is very small and the thermal velocity of these clusters is high (because of which they readily leave the beam). Therefore, the effect of ultrasmall clusters will be generally ignored. We will restrict analysis to only small clusters, which will be assumed to follow the exponential size distribution. With the effect of clusters with "magic" numbers of molecules neglected, the distribution of clusters with mean size parameter N can be represented as

$$f_N(n) = \frac{1}{N}\exp\left(\frac{-n}{N}\right). \tag{3}$$

3.1. *Transverse Component of the Thermal Velocity of Clusters*

If, when simulating cluster component S_C of the signal, we assume that the product $\zeta_n\varphi_n^2 = \zeta_N\varphi_N^2$ is the same for clusters of any size and $V_{n\perp} \equiv V_\perp(\Delta Q)$ for the nonequilibrium case, this component, in view of (1), can be written as

$$S_C = \sum_n \zeta_n \varphi_n^2 B_n n C_n = \zeta_N \varphi_N^2 B(V_\perp)\sum_n n C_n = \zeta_N \varphi_N^2 B(V_\perp)(qN_S). \tag{4}$$

When finding $B(V_\perp)$, we considered the divergent beam with its cross-sectional area in the detection zone being dependent on geometric factor R_0 (Fig. 1a). An additional broadening of the beam due to the transverse thermal motion of the particles was also taken into consideration.

The normalized particle concentration distribution over the cross section of the beam at distance ΔZ_{SD} from the skimmer at time instant $\Delta t = \Delta Z_{SD}/U$ has the form

$$B(y, V_\perp) = \frac{1}{\pi \Delta t^2 V_\perp^2} \times \left[\int_0^{2\pi}\left(\int_0^{R_0} R\exp\left(\frac{-(y^2 + R^2 - 2Ry\cos(\varphi))}{\Delta t^2 V_\perp^2}\right)dR\right)d\varphi\right]. \tag{5}$$

In particular, at $y = 0$, $B(y = 0, V_\perp)$ characterizes the thermal-diffusion-induced relative variation of the particle concentration in the central part of the beam as a function of the transverse velocity,

$$B(y = 0, V_\perp) = 1 - \exp\left(\frac{-R_0^2}{\Delta t^2 V_\perp^2}\right). \tag{6}$$

The form of this diffusion versus velocity function, $B = B(V_\perp)$, is shown in Fig. 6a, and Fig. 6b demonstrates the space distributions of $B(y, V_\perp)$ calculated for a number of velocities. The diffusion factor is seen to

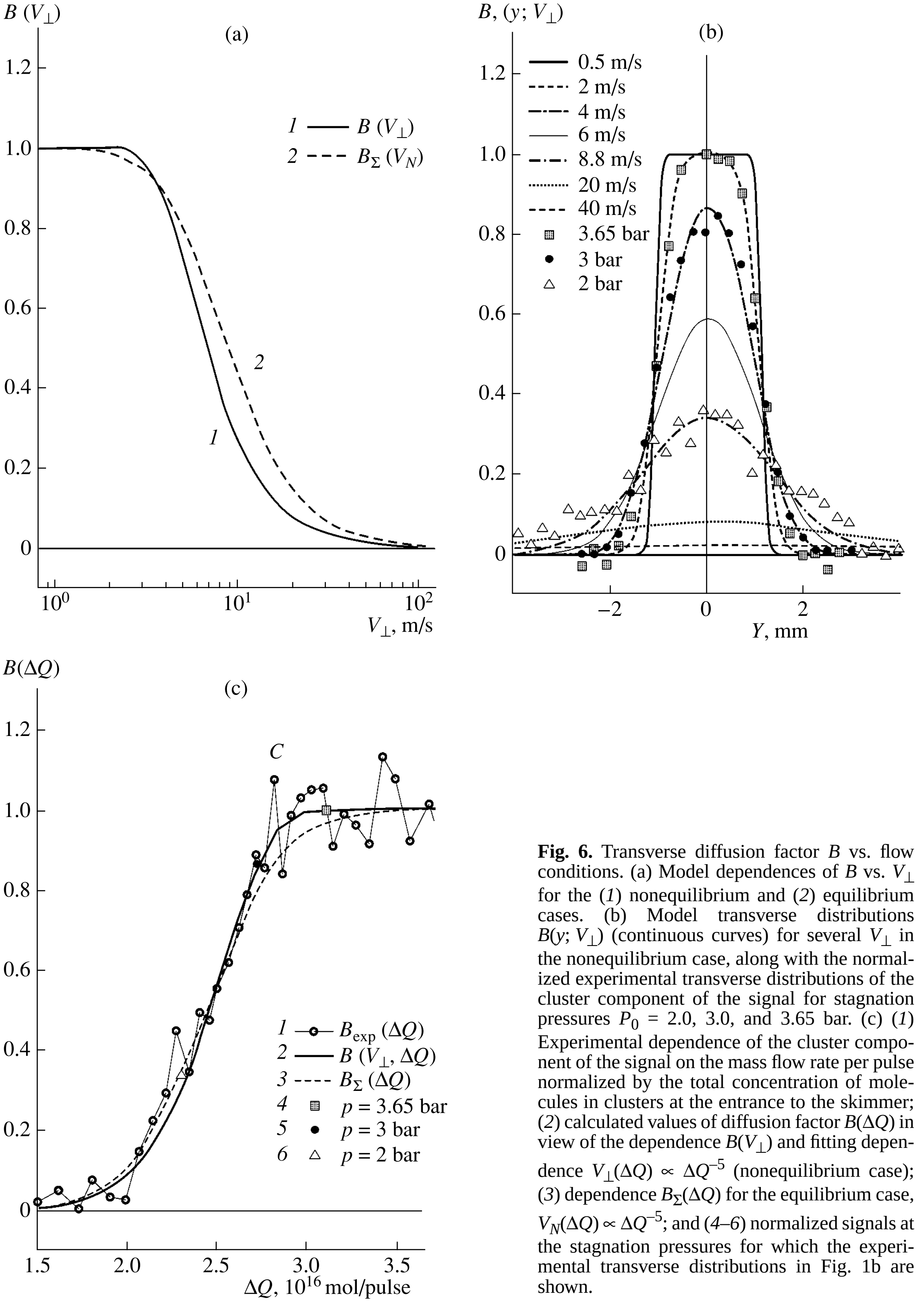


Fig. 6. Transverse diffusion factor B vs. flow conditions. (a) Model dependences of B vs. $V_\perp$ for the (*1*) nonequilibrium and (*2*) equilibrium cases. (b) Model transverse distributions $B(y; V_\perp)$ (continuous curves) for several $V_\perp$ in the nonequilibrium case, along with the normalized experimental transverse distributions of the cluster component of the signal for stagnation pressures P_0 = 2.0, 3.0, and 3.65 bar. (c) (*1*) Experimental dependence of the cluster component of the signal on the mass flow rate per pulse normalized by the total concentration of molecules in clusters at the entrance to the skimmer; (*2*) calculated values of diffusion factor $B(\Delta Q)$ in view of the dependence $B(V_\perp)$ and fitting dependence $V_\perp(\Delta Q) \propto \Delta Q^{-5}$ (nonequilibrium case); (*3*) dependence $B_\Sigma(\Delta Q)$ for the equilibrium case, $V_N(\Delta Q) \propto \Delta Q^{-5}$; and (*4*–*6*) normalized signals at the stagnation pressures for which the experimental transverse distributions in Fig. 1b are shown.

considerably influence the particle concentration in the central part of the beam only if $V_\perp > 3$ m/s. At lower velocities, its effect is negligible in our experimental conditions. Therefore, at pressures above P_C (or at corresponding mass flow rates ΔQ_C), when the thermal velocity of the particles is sufficiently low, $B(y = 0) \simeq 1$ and, accordingly, the cluster component of the signal is given by $S_C = \zeta_N \varphi_N^2 (qN_S)$. Then, based on the experimental dependence $S_C(\Delta Q)$ (see Fig. 4), we can conclude that the total concentration of the molecules in clusters near the skimmer is a quadratic function of the mass flow rate (when $\Delta Q > \Delta Q_C$),

$$(qN_S) = \frac{S_C(\Delta Q(P_0))}{\zeta_N \varphi_N^2} = \xi_C(N_S(\Delta Q) \propto (\Delta Q)^2).$$

Assuming that the molecule concentration in the clusters near the skimmer, $\zeta_C(N_S(\Delta Q))$, follows this dependence at lower mass flow rates as well and normalizing the signal associated with the cluster component by $qN_S = \zeta_C(\Delta Q)$, we can separate out the effect of the diffusion factor on the concentration of the molecules observed in the cluster component at the detector versus the mass flow rate per pulse,

$$\frac{S_C(\Delta Q)}{\zeta_N \varphi_N^2 \xi_C(\Delta Q)} = B_{\exp}(y = 0; \Delta Q). \tag{7}$$

The experimental data normalized in this way are shown in Figs. 6b and 6c. They are seen to be in good agreement with the calculated run of the particle relative concentration near the detector with regard to the beam diffusion factor versus transverse velocity $V_\perp$ (the curves $B(V_\perp)$ and $B(y, V_\perp)$). If fitting dependence $V_\perp^*(\Delta Q)$ is introduced into functional dependence $B(y = 0, V_\perp)$ (6) instead of $V_\perp$, then, fitting the calculated dependence $B(V_\perp^*(\Delta Q)$ to the experimental dependence $B_{\exp}(\Delta Q)$ yields the functional dependence $V_\perp^*(\Delta Q)$. The fitting dependence shown in Fig. 6c (curve *2*) was derived at $V_\perp^*(\Delta Q) \propto \Delta Q^{-5}$.

Note that above we did not use a specific form of the cluster size distribution in the ensemble.

In the equilibrium case, calculation of the cluster component of the signal necessitates summation of the contributions from particles of different size with regard to their spreads and specific size distributions in the cluster ensemble, $S_C = \sum_n \zeta_n \varphi_n^2 B_n(v_\perp(n)) n C_n(n)$.

For the equilibrium case, the normalized diffusion factor has the form

$$B_\Sigma(y, V_N) = \frac{\sum\limits_n B_n(y, v_n) n f_N(n)}{\sum\limits_n n f_N(n)}. \tag{8}$$

Here, the thermal velocity of the particles depends on their size, $v_n = V_N\sqrt{N/n}$, where V_N is the velocity parameter for particles with $n = N$ in a given ensemble. If the cluster size distribution is exponential, an expression for the diffusion factor in the central part of the beam can be derived analytically,

$$B_\Sigma(y = 0, V_N) = 1 - \left(\frac{\Delta t^2 V_N^2}{R_0^2 + \Delta t^2 V_N^2}\right)^2. \tag{9}$$

The dependence $B_\Sigma(V_N)$ is shown in Fig. 6a by the dashed line, and the ΔQ dependence of the diffusion factor, $B_\Sigma(\Delta Q)$, at $V_N \propto \Delta Q^{-5}$ is demonstrated in Fig. 6c (dashed line). It is seen that, early in condensation (near point *B*), the latter dependence is similar to that in the equilibrium case, while at the stage of developed condensation, the experimental dependence is closer fitted by the curve calculated for the nonequilibrium case (Figs. 6a–6c). Analysis shows that, if the effect of nonequilibrium on the cluster system is not taken into account, the ΔQ dependence of the transverse component of the thermal velocity may be approximated by the power law $V(\Delta Q) \propto \Delta Q^{-m}$. It should be noted that exponent m may vary between 4 and 5, depending on experimental and methodic errors. ΔQ dependences of $V_\perp$ for the nonequilibrium case are shown in Fig. 7 (curves *2* and *3*). It is seen that, when mass flow rate ΔQ varies in the range $(2.2–3.5) \times 10^{16}$ mol/pulse, the value of $V_\perp$ varies from 11 to 1.4 m/s.

3.2. Mean Size of Clusters

In many works devoted to clustering at different properties of the gas, nozzle designs, and flow conditions (flow temperatures and stagnation pressures), it has been found that the cluster mean size and scaling parameter Γ^* (introduced by Hagena [19]) are related by a simple power law, $N \propto (\Gamma^*)^a$. The scaling parameter characterizes flow conditions for a given nozzle design and gas properties. Note that this parameter is a linear function of the gas density in the source, n_0; hence, $N \propto (n_0)^a$. Measurements carried out by various techniques strengthened the validity of this relationship both for clusters of noble (atomic) gases (Xe_N, Kr_N [20] and Ar_N [20–22]) and for molecular gases ($(H_2O)_N$, $(NH_3)_N$ [23] and $(CO_2)_N$ [24, 25]). Experiments show that exponent $a = 2$ provides a good fit to experimental data for clusters with a small mean size ($N < 100$).

If it is assumed that the sharp increase in the signal amplitude with mass flow rate ΔQ (pressure P_0) is associated with condensation (formation of stable clusters with $n > 5$) and that stable clusters consist, on average, of 5–10 molecules [15], we obtain an estimator for cluster mean size N_B as a function of the mass flow rate per pulse in the form $N(\Delta Q) = N_B(\Delta Q/\Delta Q_B)^2$. The corresponding curve is shown by the dashed line in Fig. 4 (curve *3*, right-hand axis), where $N_B = 7.5$ for definiteness, and in Fig. 7 (curve *8*). It is seen that the cluster mean size does not exceed $N = 30–50$ molecules under the given experimental conditions and assumptions made above.

Now, with regard to the relationship $N \propto \Delta Q^2$ and the dependence $V_\perp(\Delta Q) \propto \Delta Q^{-5}$ for the nonequilibrium case, we can relate the velocity parameter and the mean size of clusters forming under different initial conditions, $V_\perp(N) \propto 1/N^{2.5}$.

3.3. Degree of Condensation

The degree of condensation in the flow near the skimmer can be found from the ratio of the monomer and cluster components of the signal (see (1)),

$$\frac{S_M}{S_C} = \frac{B_1(1-q)}{B_C q}, \quad \text{hence} \quad q = \frac{1}{1+\frac{B_C S_M}{B_1 S_C}}. \qquad (10)$$

For $\Delta Q > \Delta Q_C$, the diffusion factor for the cluster component is close to unity ($B_C \approx 1$). For the monomer component, it is found from the dependence $B_1(V_\perp)$, the corresponding transverse velocity being determined from the monomer transverse distribution in the beam. Unfortunately, the transverse distribution of monomers is wide throughout the pressure range, so that $V_\perp$ is hard to estimate with a reasonable accuracy from it. For example, at $\Delta Q = 3.4 \times 10^{16}$ mol/pulse ($P_0 = 4$ bar) for monomers, the velocity estimated from the width of the profile exceeds 40 m/s, which corresponds to $B_1 < 1/50$. At the given mass flow rate, the fraction of condensed molecules (near the skimmer) is estimated as $q < 0.16$. It is worth noting that, under these flow conditions, the mean concentration of molecules in the clusters at the beam axis near the detector is 10 times higher than the concentration of monomers ($S_C/S_M \approx 10$). Since $qN_S \propto (\Delta Q)^2$ and $N_S \propto \Delta Q$, we have $q \propto \Delta Q$ under the conditions of developed condensation. The corresponding estimator is plotted in Fig. 7 (curve *7*).

3.4. Velocity of Cluster Directed Motion and Its Spread

Velocity U of cluster directed motion and longitudinal component $V_\parallel$ of the thermal velocity of clusters in the beam were determined by comparing the experimental data with relevant calculated dependences of the modulated TOF spectrum, $S(t, y = 0)$.

For particles with given velocity parameters U and $V_\parallel$, the TOF spectrum near the detector is defined by the convolution

$$S(t, V_\parallel) = \int_{t_0}^{t} Y(t-\tau)F(\tau)d\tau, \qquad (11)$$

where $F(t)$ is the source function modulated by a rectangular pulse and

$$Y(t, U, V_\parallel) = \frac{4}{\sqrt{\pi}V_\parallel}\frac{(z_d-z_0)^3}{(t-t_0)^4} \times \exp\left(-\frac{1}{V_\parallel^2}\left(\frac{(z_d-z_0)}{(t-t_0)}-U\right)^2\right) \qquad (12)$$

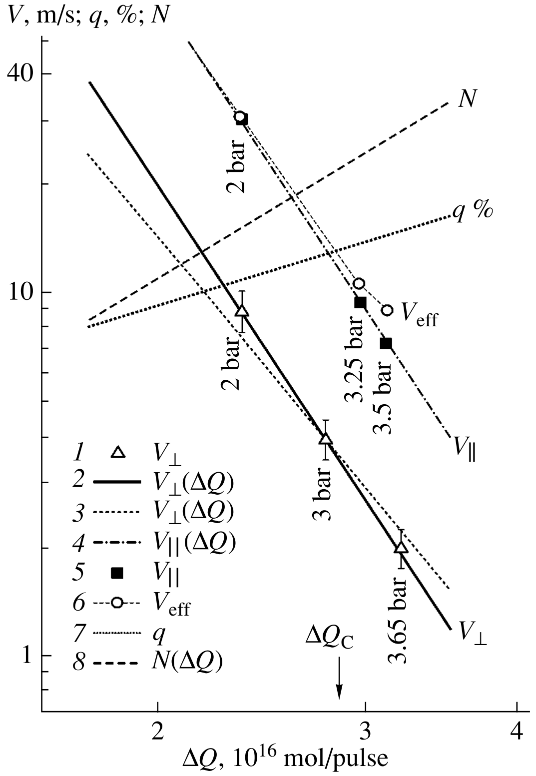


Fig. 7. Basic parameters of the cluster component of the beam vs. mass flow rate per pulse. (*1*) Experimental values of $V_\perp$ obtained from the corresponding transverse distributions for a number of stagnation pressures (see Fig. 6b); (*2*) model dependence $V_\perp(\Delta Q) \propto \Delta Q^{-5}$ obtained by fitting the calculated values of $B(\Delta Q)$ to $B_{\text{exp}}(\Delta Q)$; (*3*) the dependence $V_\perp \propto \Delta Q^{-4}$ shown for comparison; (*4*) the dependence of the longitudinal component of the cluster's thermal velocity, $V_\parallel(\Delta Q)$; (*5*) values of $V_\parallel$ corresponding to the experimental conditions in Fig. 5; (*6*) values of effective velocities V_{eff} formally derived from the front width (Fig. 5) at $\gamma_C = 0$; (*7*) degree of condensation q of molecules in the jet near the skimmer; and (*8*) estimated dependence of cluster mean size N in the jet on the mass flow rate per pulse.

is the response of the "volume" detector (with coordinate z_d) to the action of a source of type $\delta(z_0, t_0)$ [26] (the solution is given in the normalized form).

To correctly calculate the TOF spectrum, it is necessary to sum the contributions from clusters of different size with regard to the density of cluster size distribution $f(n)$, size-dependent mean velocity U_n of directed motion, and variance $V_\parallel$ of the directed motion velocity. Then, the normalized signal takes the form

$$S_\Sigma(t) = \frac{\sum_n nf(n)\int_{t_0}^{t} Y_n(U_n, V_{\parallel n}, (t-\tau)F(\tau)d\tau)}{\sum_n nf(n)}. \qquad (13)$$

Below, we will assume that, for an ensemble of small clusters, the density of their concentration distribution over sizes, $f_N(n)$, varies exponentially with parameter N depending on flow conditions, $N(\Delta Q)$. To simplify calculations, our consideration will be restricted to the nonequilibrium case, when the variance of velocities in the cluster ensemble is characterized by single parameter $V_{N\|}(\Delta Q)$, which depends on flow conditions alone. Let the dependence of the directed motion velocity on the cluster size, $U(n)$, be similar to that obtained in [18] for small water clusters. From experimental data in [18], it follows that this dependence can be approximated as

$$U(n) = U_C\left(1 - \gamma_C\frac{n - n_C}{n + n_C}\right), \tag{14}$$

where U_C and γ_C may generally depend on flow conditions and $n_C = 6$.

Fragments of the TOF spectrum near the front of the step mark are shown in Fig. 5 for $P_0 = 2.0$, 3.25, and 3.50 bar). As the pressure rises from 2.0 to 3.5 bar, the directed motion velocity determined from the time of arrival of the half-height of the front mark insignificantly drops from 509.5 to 505.6 m/s. A rise in the pressure is accompanied by narrowing of the front; that is, the variance of the directed motion velocity of clusters in the beam decreases. Note that the diffusion of the front in solution (13) is due both to the longitudinal component of the random motion velocity of clusters, $V_\|$, and to the dependence $U = U(n; U_C, \gamma_C, n_C)$. Therefore, at given distribution $f_N(n)$, the mark diffuses even if $V_\| = 0$. The contribution of both factors can be approximately taken into account through $V_{\text{eff}} \approx \sqrt{V_\|^2 + \Delta_\|^2}$, where V_{eff} is the formally introduced velocity describing the width of the mark front that includes the variance of the longitudinal velocity. In simulation, additional contribution $\Delta_\|$ depends on parameter γ_C. This contribution noticeably distorts the shape of the front, making it asymmetric. When the experimental data are fitted by the model spectrum, the relative contribution of either factor to the front diffusion depends on a set of interrelated adjusting parameters. As applied to the model approximations used, one can choose several sets of these parameters that provide a reasonable fit to TOF spectra found experimentally (within the experimental errors). For the calculated spectra shown in Fig. 5, the mean sizes of clusters ($N = 15$, 24, and 26.5 for pressures $P_0 = 2.0$, 3.25, and 3.5 bar, respectively) are taken according to the estimator $N(\Delta Q)$ obtained earlier with regard for the dependence $\Delta Q(P_0)$. The calculated front of the step was fitted to its experimental shape by varying γ_C. It was found that $\gamma_C = 0.04$ gives the best fit to each of the spectra taken in the given pressure range. The corresponding (fitting) values of the thermal velocities ($V_\| = 30.4$, 9.4, and 7.3 m/s) are related to the cluster mean size as $V_\| \propto N^{-2.5}$ (or $V_\| \propto \Delta Q^{-5}$). The same relationship is true for $V_\perp$. The ΔQ dependences of both velocities are shown in Fig. 7, which also plots effective velocity V_{eff} that was formally obtained from the front width simulated at $\gamma_C = 0$. Since $V_{\text{eff}} \approx \sqrt{V_\|^2 + \Delta_\|^2}$, the additional velocity-related contribution is estimated as $\Delta_\| \approx 5.3$ m/s (at $P_0 = 3.5$ bar, $V_\| = 7.3$ m/s, and $V_{\text{eff}} = 9$ m/s). This contribution is seen to become comparable with the contribution from the random motion velocity with rising pressure.

Supposing that the relationship between velocities $V_\|$ and $V_\perp$ that is observed in Fig. 7 is explained by a decrease in $V_\perp$ because of the geometrical cooling factor, which influences the process over the distance between the point of freezing of translational degrees of freedom (the point of transition to collisionless flow) and the skimmer, one can locate this point (it is ≈11 mm distant from the nozzle or ≈27.5 mm from the skimmer). Hence, the conditions favorable for jet particle condensation last $\Delta t_C \approx 22$ μs).

4. CONCLUSIONS

Thus, we studied the clustering of CF_2HCl molecules flowing out of a supersonic nozzle. Using original cluster beam diagnostics methods based on UV multiphoton ionization, IR photodissociation, and TOF mass spectrometry, as well as the model presented in this work, the basic characteristics of $(CF_2HCl)_n$ cluster beams were determined.

It is shown that, in the normal operating mode of the nozzle (pure Freon-22 without a carrier gas), developed condensation sets even if the stagnation pressure is as low as $P_0 = 1.25$ bar. Under the conditions of developed condensation, the total concentration of molecules entering into the clusters of the jet is $qN_S \propto \Delta Q^2$.

It is found that the variation of the cluster component intensity in the beam near the detector with mass flow rate per pulse, $S_C(\Delta Q)$, is associated both with a change in the fraction of molecules in the condensed phase of the jet, $q(\Delta Q) \propto \Delta Q$, and with the diffusion of the cluster beam in the transverse direction. The amount of diffusion depends on the transverse component of the random motion velocity of clusters in the beam. If the effect of equilibrium conditions (which are assumed in simulation) in the cluster system is ignored, this velocity as a function of ΔQ can be approximated as $V(\Delta Q) \propto \Delta Q^{-m}$, where exponent m may vary between 4 and 5 with regard to experimental and methodic errors. The best agreement between the measured and calculated dependences is observed on the assumption that an ensemble of small clusters is nonequilibrium. Under these conditions, the random motion velocity of the clusters is characterized by single cluster-size-independent parameter $V_n = V^*(\Delta Q)$ (and $V(\Delta Q) \propto \Delta Q^{-5}$).

It is found that the velocity of directed motion of small clusters slightly drops (from 509.5 to 505.6 m/s) as pressure P_0 rises from 2.0 to 3.5 bar. However, longitudinal component $V_{||}$ of the thermal velocity drops quickly from 30.4 to 7.3 m/s as P_0 rises from 2.0 to 3.5 bar. This velocity varies with the mass flow rate as $V_{||}(\Delta Q) \propto \Delta Q^{-5}$.

Translational velocity $U(n)$ depends on the cluster size within a cluster ensemble. This fact causes an additional asymmetric broadening of the mark front in TOF measurements.

Mean cluster size N and degree of clustering q in the jet are estimated. Depending on the pressure in the range $1.5 < P_0 < 4$ bar, N varies from 10 to 30 and q does not exceed 0.16 at $P_0 = 4$ bar.

ACKNOWLEDGMENTS

The authors thank G.N. Makarov for fruitful discussions.

This work was supported by the Russian Foundation for Basic Research, grant no. 03-02-17080.

REFERENCES

1. V. Yu. Baranov, A. P. Dyad'kin, D. D. Malyuta, *et al.*, Proc. SPIE **4165**, 314 (2000).
2. R. V. Ambartsumyan, Yu. A. Gorokhov, V. S. Letokhov, and G. N. Makarov, Zh. Éksp. Teor. Fiz. **69**, 1956 (1975) [Sov. Phys. JETP **42**, 993 (1975)].
3. R. Y. Jensen, O. F. Judd, and Y. A. Sullivan, Los Alamos Sci. **3**, 2 (1982).
4. S. S. Alimpiev, G. S. Baronov, S. M. Karavaev, *et al.*, Kvantovaya Élektron. (Moscow) **10**, 376 (1983).
5. G. N. Makarov, D. E. Malinovskii, and D. D. Ogurok, Zh. Tekh. Fiz. **69** (1), 35 (1999) [Tech. Phys. **44**, 31 (1999)].
6. G. N. Makarov, V. N. Lokhman, D. E. Malinovskii, *et al.*, Khim. Fiz. **18** (3), 71 (1999).
7. V. N. Lokhman, D. D. Ogurok, and E. A. Ryabov, Chem. Phys. **271**, 357 (2001).
8. *Atomic and Molecular Beam Methods*, Ed. by G. Scoles (Oxford Univ. Press, New York, 1988–1992), Vols. 1–2.
9. H. Pauly, *Atom, Molecule and Cluster Beams* (Springer Verlag, New York, 2000), Vol. 2.
10. V. N. Lokhman, E. A. Ryabov, and D. D. Ogurok, Pis'ma Zh. Tekh. Fiz. **30** (8), 88 (2004) [Tech. Phys. Lett. **30**, 345 (2004)].
11. O. F. Hagena and W. Obert, J. Chem. Phys. **56**, 1793 (1972).
12. A. W. Castleman and K. H. Bowen, J. Phys. Chem. **100**, 12911 (1996).
13. B. M. Smirnov, Usp. Fiz. Nauk **173**, 609 (2003) [Phys. Usp. **46**, 589 (2003)].
14. A. V. Dem'yanenko, V. N. Lokhman, D. D. Ogurok, *et al.*, Chem. Phys. Lett. **320**, 594 (2000).
15. *Physics of Clusters: A Collection of Scientific Works*, Ed. by A. A. Vostrikov and A. K. Rebrov (Izd. SO AN SSSR, Novosibirsk, 1987) [in Russian].
16. A. A. Vostrikov, A. K. Rebrov, and B. E. Semyachkin, Zh. Tekh. Fiz. **50**, 2425 (1980) [Sov. Phys. Tech. Phys. **25**, 1418 (1980)].
17. A. A. Vostrikov, S. G. Mironov, and B. E. Semyachkin, Zh. Tekh. Fiz. **53**, 81 (1983) [Sov. Phys. Tech. Phys. **28**, 46 (1983)].
18. D. Dreyfuss and H. Y. Wachman, J. Chem. Phys. **76**, 2031 (1982).
19. O. F. Hagena, Surf. Sci. **106**, 101 (1981).
20. R. Karnbach, M. Joppien, J. Stapelfeldt, *et al.*, Rev. Sci. Instrum. **64**, 2838 (1993).
21. O. F. Hagena, Z. Phys. D **4**, 291 (1987).
22. U. Buck and R. Krohne, J. Chem. Phys. **105**, 5408 (1996).
23. C. Bobbert, S. Schütte, C. Steinbach, and U. Buck, Eur. Phys. J. D **19**, 183 (2002).
24. G. D. Stein and J. A. Armstrong, J. Chem. Phys. **58**, 1999 (1973).
25. G. Torchet, H. Bouchier, J. Farges, *et al.*, J. Chem. Phys. **81**, 2137 (1984).
26. B. N. Eldridge and M. L. Yu, Rev. Sci. Instrum. **58**, 1014 (1987).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 856–863. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 40–47.
Original Russian Text Copyright © 2005 by Shiryaeva, Grigor'ev, Volkova.

GASES AND LIQUIDS

On the Possibility of Corona Initiation Near a Conducting Drop Nonlinearly Vibrating in an External Electrostatic Field

S. O. Shiryaeva, A. I. Grigor'ev, and M. V. Volkova
Demidov State University, Sovetskaya ul. 14, Yaroslavl, 150000 Russia
e-mail: grig@uniyar.ac.ru

Received December 1, 2004

Abstract—An analytical asymptotic expression for the field strength near an ideal incompressible electrically conducting liquid drop nonlinearly vibrating in external electrostatic field E_0 is found in an order of 5/2 in a small parameter. The small parameter here is the amplitude of deformation of the spherical shape of the drop. It is found that the strength of the electric field resulting at the tops of the drop exceeds the corona-initiating field even if E_0 is one order of magnitude lower than the value at which the drop becomes unstable against the induced charge (that is, at such values of E_0 as are observed in storm clouds in full-scale experiments). © *2005 Pleiades Publishing, Inc.*

(1) According to one of the existing models, a lightning discharge in storm clouds is triggered by a corona near a large drop [1–3]. It therefore seems to be of interest to calculate the electric field strength near a charged water drop nonlinearly vibrating in an external uniform electrostatic field. In such a general statement, the problem is too complicated to be treated analytically; so, it is reasonable to split it into subproblems that are much simpler for nonlinear analysis. For example, in [3], we analyzed the possibility of corona initiation near a charged nonlinearly vibrating spherical drop and showed that such a phenomenon may take place when the charge of the drop is sufficiently high but still is three times as small as a value critical in terms of Rayleigh instability. Yet, the problem persists, since charges close to one-third of the Rayleigh critical value are rarely encountered in cloud drops in full-scale measurements [4]. For a drop in a uniform external electrostatic field, such calculations turn out to be much more complicated, because the equilibrium shape of the drop under these conditions becomes nearly spheroidal rather than being spherical. This circumstance has impeded investigations into nonlinear vibration of the drop in external fields, although nonlinear vibration of spherical charged drops has been studied at length for the two last decades (see, for example, [5–8] and Refs. therein). When asymptotic calculations of the vibration of the drop are conducted in the spherical coordinate system, the spheroidicity of its equilibrium shape in a uniform electrostatic field gives rise to two small parameters: the amplitude of the equilibrium spheroidal deformation and the amplitude of nonlinear vibration (in the spheroidal coordinate system, the calculations are even more awkward and less illustrative). Note that calculations of spheroidal drop vibration in a uniform electrostatic field have been performed to date only in an approximation linear in vibration amplitude [9–14].

As a dimensionless parameter characterizing the equilibrium spheroidal deformation of a conducting drop in an electrostatic field, it is natural to take its eccentricity e. The ratio ξ/R (where ξ is the amplitude of capillary vibrations of a spheroidal drop and R is the radius of the equivalent sphere) is the second small parameter, $\varepsilon \equiv \xi/R$, which has the meaning of the dimensionless vibration amplitude. In practice, when calculating vibrations of a spheroidal drop in the first order of smallness in amplitude ε, one should also take into account the second order of smallness in eccentricity e (i.e., in electrostatic field dimensionless strength) [9–14]. In other words, the problem becomes in a sense nonlinear when expansions in both small parameters and terms $\sim\varepsilon \cdot e^2$ are left in the solution.

In asymptotic calculations using expansions in several small parameters, it is necessary to correlate their orders of smallness. For example, if, in the problem considered, it is assumed that $\varepsilon \sim e^2$ (see [11–13]), calculations made in the approximation $\sim\varepsilon \cdot e^2$ should also involve terms $\sim\varepsilon^2$. This was not done in [11–13], so that the results obtained in those works are incorrect.

In this paper, we analyze nonlinear vibrations of an uncharged drop in a uniform electrostatic field, assuming that $\varepsilon \sim e^2$ and leaving terms $\sim\varepsilon \cdot e^2$ and $\sim\varepsilon^2$. The strength of the electrostatic field near the surface of the drop is calculated accurate to $\sim\varepsilon^{5/2}$, which is necessary for due regard to interaction between the stationary spheroidal deformation and vibration modes.

(2) Consider the evolution of the surface shape of an uncharged drop nonlinearly vibrating in uniform electrostatic field $\mathbf{E}_0$. We assume that the ideal incompress-

ible perfectly conducting liquid has density ρ and surface tension coefficient σ.

In the absence of an external field, the drop is an equilibrium sphere of radius R. Weak external uniform electrostatic field $\mathbf{E}_0$ causes a small distortion of the spherical shape of the drop: it elongates in the field direction, taking the form a prolate spheroid up to terms proportional to $e^2 = (9E_0^2R/16\pi\sigma)$. The equation of the generatrix of the surface of such an equilibrium drop is written in the form

$$r(\theta) \equiv 1 + \frac{1}{3}e^2P_2(\mu); \quad \mu \equiv \cos\theta, \tag{1}$$

where $P_2(\mu)$ is the Legendre polynomial.

Dimensionless quantity $W_e \equiv E_0^2R/\sigma$ is the so-called Taylor parameter, which characterizes the stability of the drop against the surface charge induced by external uniform electrostatic field E_0. The critical (in terms of instability) value of this parameter is $W_E \equiv W_E^* \approx 2.62$ [10].

Let, at initial time instant $t = 0$, an equilibrium weakly spheroidal drop with eccentricity e and the generatrix defined by (1) experience axisymmetric perturbation $\xi(\theta, t)$ of fixed amplitude ξ, which is much smaller than the radius of the drop $(\xi/R) \equiv \varepsilon \ll 1$. The perturbation makes the drop vibrate about its equilibrium spheroidal shape. Our aim is to find the shape of the drop's generatrix as a function of time (i.e., at $t > 0$). In the dimensionless variables such that $\rho = \sigma = R = 1$ (we retain the previous notation), the equation of the generatrix in the spherical coordinate system with the origin at the center of mass of the drop has the form

$$r(\theta, t) \equiv r(\theta) + \xi(\theta, t)$$

$$\equiv 1 + \frac{1}{3}e^2P_2(\mu) + \xi(\theta, t) \equiv 1 + f(\theta, t); \tag{2}$$

$$e^2 = \frac{9E_0^2}{16\pi}; \quad |\xi| \ll 1; \quad |f| \ll 1.$$

The liquid flow in the drop that is caused by the distortion of its equilibrium shape is assumed to be potential with velocity field potential $\psi(\mathbf{r}, t)$ and velocity field $\mathbf{V}(\mathbf{r}, t) = \operatorname{grad}(\psi(\mathbf{r}, t))$. Taking into account that the velocities of gasdynamic liquid flows in the drop are much lower than the rate of propagation of electromagnetic interactions, we may assume the electric field near the drop to be electrostatic and describe it through potential $\Phi(\mathbf{r}, t)$, which is related to field strength $\mathbf{E}$ by the well-known expression $\mathbf{E} = -\operatorname{grad}(\Phi)$.

Mathematically, the problem of analytical asymptotic calculation of nonlinear vibrations of a conducting liquid drop in an external uniform electrostatic field is stated as

$$\Delta\psi(\mathbf{r}, t) = 0; \quad \Delta\Phi(\mathbf{r}, t) = 0; \tag{3}$$

$$r \longrightarrow 0; \quad \psi(\mathbf{r}, t) \longrightarrow 0; \tag{4}$$

$$r \longrightarrow \infty; \quad \Phi(\mathbf{r}, t) \longrightarrow -E_0r\cos(\theta); \tag{5}$$

$$\begin{gathered} r = r(\theta) + \xi(\theta, t)\colon \frac{\partial\xi}{\partial t} = \frac{\partial\psi}{\partial r} - \frac{1}{r^2}\frac{\partial\xi}{\partial\theta}\frac{\partial\psi}{\partial\theta}; \\ \Delta p - \frac{\partial\psi}{\partial t} - \frac{1}{2}(\nabla\psi)^2 + p_E = p_\sigma; \end{gathered} \tag{6}$$

$$\Phi(\mathbf{r}, t) = 0; \tag{7}$$

$$\begin{gathered} \int_V r^2 dr\sin\theta d\theta d\varphi = \frac{4}{3\pi}, \\ V = [0 \le r \le r(\theta) + \xi(\theta, t),\ 0 \le \theta \le \pi,\ 0 \le \varphi \le 2\pi]; \end{gathered} \tag{8}$$

$$\int_V \mathbf{e}_r \cdot r^3 dr\sin\theta d\theta d\varphi = 0; \tag{9}$$

$$\begin{gathered} \oint_S (\mathbf{n} \bullet \nabla\Phi)dS = 0, \\ S = [r = r(\theta) + \xi(\theta, t),\ 0 \le \theta \le \pi,\ 0 \le \varphi \le 2\pi]; \end{gathered} \tag{10}$$

$$\begin{gathered} t = 0\colon \xi(\theta, t) = \xi_0P_0(\mu) + \xi_1P_1(\mu) + \varepsilon\sum_{i \in \Xi} h_iP_i(\mu), \\ \frac{\partial\xi(\theta, t)}{\partial t} = 0; \quad \sum_{i \in \Xi} h_i = 1 \end{gathered} \tag{11}$$

(where Δ is the Laplacian).

Conditions (8) and (9) must be fulfilled at any time instant $t \ge 0$. Therefore, at $t = 0$, they specify the amplitudes of the zeroth, ξ_0, and first, ξ_1, modes in the expansion of initial perturbation $\xi(\theta)$ in Legendre polynomials. In other words, the amplitudes of the zeroth and first modes cannot be taken arbitrary and are specified by the form of the initial deformation.

In expressions (6)–(11), Δp is the difference between constant pressures inside and outside the drop in the equilibrium state;

$$p_E = \frac{1}{8\pi}(\nabla\Phi)^2$$

is the electric field pressure; $p_\sigma = \operatorname{div}_S\mathbf{n}$ is the surface tension force pressure (div_S is the surface divergence); $\mathbf{n}$ is the unit vector normal to surface (2); ε is the dimensionless amplitude of the initial perturbation of the shape of the drop (one of the small parameters of the problem); h_i is the coefficient specifying the partial contribution of an ith vibration mode to the total initial perturbation; Ξ is the set of the numbers of initially excited vibration modes;

$$\xi_0 \approx -\sum_{i \in \Xi}\left[\varepsilon^2 \frac{h_i^2}{(2i+1)} + \varepsilon e^2 \frac{2}{15} h_i \delta_{i,2}\right] + O(\varepsilon^3),$$

$$\xi_1 \approx -\sum_{i \in \Xi}\left[\varepsilon^2 \frac{9 i h_{i-1} h_i}{(2i-1)(2i+1)} + \varepsilon e^2 \frac{9}{35} h_i \delta_{i,3}\right] + O(\varepsilon^3), \qquad (12)$$

are constants that are found from conditions (8) and (9) at the initial time instant accurate to terms of order of smallness $\sim\varepsilon \cdot e^2$ and $\sim\varepsilon^2$; and $\delta_{i,j}$ is the Kronecker delta.

To simplify subsequent expansions and have a chance to easily separate out terms due to the spheroidicity of the drop in the final expressions, we introduce formal parameter β such that $e^2 \approx \beta\varepsilon$. Since it was assumed that $e^2 \sim \varepsilon$, it is obvious that $\beta \sim 1$. The product $\beta\varepsilon$ can then be replaced with e^2. It is also remembered that, when seeking for the equilibrium shape of the drop in the field, we obtained a relationship between the eccentricity and electric field strength that in the dimensionless variables has the form

$$e^2 = \frac{9}{16\pi} E_0^2.$$

(3) To solve the problem considered, we will use the method of many scales (see also [3, 5–8]). Desired functions $\xi(\theta, t)$, $\psi(\mathbf{r}, t)$, and $\Phi(\mathbf{r}, t)$ will be represented in the form of expansions in powers of small parameter ε under the assumption that these functions depend not merely on time t but on various time scales defined via small parameter $p\varepsilon$, $T_m \equiv \varepsilon^m t$ ($m = 0, 1, 2, \ldots$). Our analysis will be restricted to the quadratic approximation, allowing us to find the dependences of the desired quantities on only two time scales T_0 and T_1,

$$\xi(\theta, t) = \varepsilon\xi^{(1)}(\theta, T_0, T_1) + \varepsilon^2\xi^{(2)}(\theta, T_0) + O(\varepsilon^3);$$

$$\psi(\mathbf{r}, t) = \varepsilon\psi^{(1)}(\theta, T_0, T_1) + \varepsilon^2\psi^{(2)}(\theta, T_0) + O(\varepsilon^3);$$

$$\Phi(\mathbf{r}, t) = \varepsilon^{1/2}(\Phi^{(0)}(r, \theta) + \varepsilon\Phi^{(1)}(r, \theta, T_0, T_1) \qquad (13)$$

$$+ \varepsilon^2\Phi^{(2)}(r, \theta, T_0)) + O(\varepsilon^{7/2}).$$

The electric field potential is expanded in half-integer powers of parameter ε, since $\varepsilon \sim E_0^2$; consequently, $\varepsilon^{1/2} \sim E_0$. Potential component $\varepsilon^{1/2}\Phi^{(0)}(r, \theta)$ is determined near the equilibrium weakly spheroidal shape of the drop,

$$\varepsilon^{1/2}\Phi^{(0)}(r, \theta) = \left(\frac{16\pi}{9}\varepsilon\beta\right)^{1/2} \frac{1}{r^2}(1 - r^3)\cos\theta$$

$$\equiv E_0 \frac{1}{r^2}(1 - r^3)\cos\theta,$$

and corrections $\Phi^{(1)}(r, \theta, T_0, T_1)$ and $\Phi^{(2)}(r, \theta, T_0)$ are related to the distortion of the equilibrium shape.

For the pressures of the electric field, p_E, and surface tension forces, p_σ, entering into dynamic boundary condition (6), the expansions are as follows:

$$p_E = p_E^{(\mathrm{eq})} + \varepsilon^2 p_E^{(2)}(\xi) + O(\varepsilon^3);$$

$$p_\sigma = p_\sigma^{(\mathrm{eq})} + \varepsilon p_\sigma^{(1)}(\xi) + \varepsilon^2 p_\sigma^{(2)}(\xi) + O(\varepsilon^3), \qquad (14)$$

where components

$$p_E^{(\mathrm{eq})} = \frac{\varepsilon}{8\pi}(\nabla\Phi^{(0)})^2\Big|_{r = r(\theta)}$$

and

$$p_\sigma^{(\mathrm{eq})} = \mathrm{div}_S \overline{\mathbf{n}^{(\mathrm{eq})}}\Big|_{r = r(\theta)}$$

are independent of perturbation $\xi(\theta, t)$. These are the pressures at the surface of the equilibrium spheroidal drop, i.e., satisfy the pressure balance

$$r = r(\theta){:}\ \Delta p + p_E^{(\mathrm{eq})} = p_\sigma^{(\mathrm{eq})}$$

and, consequently, compensate one another.

Substituting expansions (13) and (14) into system (3)–(11), one can obtain a set of boundary-value problems of various orders of smallness for functions $\xi^{(m)}$, $\psi^{(m)}$, and $\Phi^{(m)}$ ($m = 1, 2$). It is obvious that either of functions $\psi^{(m)}$ and $\Phi^{(m)}$ must satisfy linear equations (3).

Solutions to Eqs. (3) for the first- and second-order functions that meet conditions (4) and (5) are written the form

$$\psi^{(m)}(r, \theta, T_0, T_1)$$

$$= \sum_{n=1}^{\infty} D_n^{(m)}(T_0, T_1) r^{(n)} P_n(\mu) \quad (m = 1, 2);$$

$$\Phi^{(m)}(r, \theta, T_0) = \sum_{n=0}^{\infty} F_n^{(m)}(T_0) r^{-(n+1)} P_n(\mu).$$

Corrections to the equilibrium shape of the drop are also represented in the form of expansions in Legendre polynomials,

$$\xi^{(m)}(\theta, T_0, T_1) = \sum_{n=0}^{\infty} M_n^{(m)}(T_0, T_1) P_n(\mu) \quad (m = 1, 2).$$

(4) Eventually, we arrive at the following analytical expression for the shape of an uncharged drop vibrating in an external uniform electric field:

$$r(\theta, t) = 1 + \frac{1}{3}e^2 P_2(\mu) + \varepsilon\sum_{n=0}^{\infty} M_n^{(1)}(t) P_n(\mu)$$

$$+ \varepsilon^2 \sum_{n=0}^{\infty} M_n^{(2)}(t) P_n(\mu) + O(\varepsilon^3),$$

where

$$M_n^{(1)}(t) = \delta_{n,i} h_i \cos\left[\left(\omega_n - \frac{e^2}{2\omega_n} G1(n)\right)t\right];$$

$$M_0^{(2)}(t) = -\sum_{i \in \Xi}\left[\frac{2}{15}\beta\delta_{i,2} h_i \cos(\omega_i t) + \frac{h_i^2}{(2i+1)}(\cos(\omega_i t))^2\right];$$

$$M_1^{(2)}(t) = -\sum_{i \in \Xi}\left[\frac{9}{35}\beta\delta_{i,3} h_i \cos(\omega_i t) + \frac{9ih_{i-1}h_i}{(2i-1)(2i+1)}\cos(\omega_{i-1}t)\cos(\omega_i t)\right] \ (\forall n \ge 2);$$

$$M_n^{(2)}(t) = \sum_{i \in \Xi} h_i \beta\{\delta_{i,n+2}\mu_{n+2}[\cos(\omega_{n+2}t) - \cos(\omega_n t)] + \delta_{i,n-2}\mu_{n-2}[\cos(\omega_{n-2}t) - \cos(\omega_n t)]\}$$
$$+ \frac{1}{2}\sum_{i \in \Xi}\sum_{j \in \Xi} h_i h_j\{\lambda_{ijn}^{(+)}[\cos((\omega_i + \omega_j)t) - \cos(\omega_n t)] + \lambda_{ijn}^{(-)}[\cos((\omega_i - \omega_j)t) - \cos(\omega_n t)]\};$$

$$\mu_{n+2} = \frac{G2(n)}{(\omega_n^2 - \omega_{n+2}^2)}; \quad \mu_{n-2} = \frac{G3(n)}{(\omega_n^2 - \omega_{n-2}^2)};$$

$$\lambda_{lmn}^{(\pm)} = \frac{(\gamma_{lmn} \pm \omega_l\omega_m\eta_{lmn})}{[\omega_n^2 - (\omega_l \pm \omega_m)^2]}; \qquad (15)$$

$$G1(n) \equiv \frac{n}{3}[(3n(n+1) - 10)K_{2,n,n} + 12(nK_{1,n-1,n}K_{1,n,n-1} + (n+2)K_{1,n+1,n}K_{1,n,n+1}) - 8];$$

$$G2(n) \equiv \frac{1}{3}[(n^3 + 3n^2 - 10n - 8)K_{2,n+2,n} + 12n(n+2)K_{1,n+1,n}K_{1,n+2,n+1}];$$

$$G3(n) \equiv \frac{n}{3}[(5n^2 - 21n + 14)K_{2,n-2,n} + 12nK_{1,n-1,n}K_{1,n-2,n-1}];$$

$$\gamma_{lmn} = K_{lmn}[\omega_n^2(n - m + 1) + 2n(m(m+1) - 1)] + \alpha_{lmn}\frac{1}{m}\omega_m^2;$$

$$\eta_{lmn} = K_{lmn}\left(\frac{n}{2} - m + 1\right) + \alpha_{lmn}\frac{1}{m}\left(1 + \frac{n}{2l}\right);$$

$$\omega_n^2 = n(n-1)(n+2);$$

$$\alpha_{lmn} = -\sqrt{l(l+1)m(m+1)}\,C_{l0m0}^{n0}C_{l(-1)m1}^{n0};$$

$$K_{lmn} = [C_{l0m0}^{n0}]^2;$$

and C_{l0m0}^{n0} and $C_{l(-1)m1}^{n0}$ are the Clebsch–Gordan coefficients, and $\delta_{m,n}$ is the Kronecker delta.

Remarkably, in these relationships, the positions of internal nonlinear resonances (taking place when the denominators in coefficients $\lambda_{lmn}^{(\pm)}$ vanish) are independent of the spheroidicity (eccentricity) of a drop having the equilibrium shape of a prolate spheroid. Thus, there arises the need to carry out calculations in the next order of smallness for the model considered to approach to reality.

(5) Note that correction $\Phi^{(1)}$ alone will suffice to determine the shape of a nonlinearly vibrating drop up to terms of order $\sim\varepsilon^2$ in expansion (13) for the electrostatic potential, since the electrostatic field influences the shape only via pressure p_E and the addition to this pressure due to perturbation $\xi(\theta, t)$ is quadratic in amplitude ε of this perturbation and has the form

$$p_E^{(2)}(\xi) = \frac{1}{8\pi}\left[\frac{\partial}{\partial r}((\nabla\Phi^{(0)})^2)\xi^{(1)} + 2\nabla\Phi^{(0)}\cdot\nabla\Phi^{(1)}\right],$$

that is, depends only on $\Phi^{(0)}$ and $\Phi^{(1)}$.

Now our goal is to calculate electrostatic field strength $\mathbf{E} = -\mathrm{grad}(\Phi)$ at the surface of the nonlinearly vibrating drop. It is easy to check that the expression for the magnitude of strength E written with allowance for terms of order $\sim\varepsilon^{3/2}$ inclusive misses the most intriguing effect, namely, interaction of the equilibrium spheroidal shape of the drop with vibration-induced perturbation $\xi(\theta, t)$ of the shape. Indeed, in this approximation, the expression for the field strength near the surface,

$$E \approx -\varepsilon^{1/2}\left[\frac{\partial\Phi^{(0)}}{\partial r} + \varepsilon\left(\frac{\partial\Phi^{(1)}}{\partial r} - \frac{1}{r}\frac{\partial\Phi^{(0)}}{\partial\theta}\left(\frac{\partial\xi^{(1)}}{\partial\theta} + \frac{\beta}{3}\frac{\partial P_2(\mu)}{\partial\theta}\right)\right)\right] + O(\varepsilon^{5/2}),$$

does not involve terms proportional to the product of the spheroidal deformation and the amplitude of the surface profile perturbation.

To take into account the effect of interest, it is necessary to leave terms of order $\sim\varepsilon^{5/2}$ in the expression for field strength E. This, in turn, requires correction $\Phi^{(2)}$ in expansion (13) for the electrostatic potential to be calculated.

(6) The closed electrostatic problem consists of Eq. (3) for potential $\Phi(\mathbf{r}, t)$, as well as boundary and additional conditions (5), (7), and (10). A solution is sought in the form of expansion (13). Substitution of (13) into (3), (5), (7), and (10) splits the general problem into partial problems of various orders of smallness the solution of which allows one to find corrections $\Phi^{(1)}(\mathbf{r}, t)$ and $\Phi^{(2)}(\mathbf{r}, t)$ to the potential.

It is remembered that, by definition, component $\varepsilon^{1/2}\Phi^{(0)}(r, \theta)$ of the potential describes the field near an uncharged equilibrium weakly spheroidal drop placed in an external uniform field. An exact expression for the potential of such a field was derived in [15]. In the spherical coordinates, expansion of this expression in powers of eccentricity e (and, consequently, in powers of parameter ε, since $e^2 = \beta\varepsilon$) has the form

$$\Phi^{(0)}(r, \theta) \approx \left(\frac{16\pi}{9}\beta\right)^{1/2}$$

$$\times[\Phi^{(0,0)}(r, \theta) + \varepsilon\beta\Phi^{(0,1)}(r, \theta) + O(\varepsilon^2)];$$

$$\Phi^{(0,0)}(r, \theta) \equiv \frac{1}{r^2}(1 - r^3)\cos(\theta);$$

$$\Phi^{(0,1)}(r, \theta) \equiv \frac{1}{20r^4}(8r^2 - 3 + 15\cos(2\theta))\cos(\theta).$$

The potential of the electrostatic field near a nonlinearly vibrating weakly spheroidal drop is sought in the spherical coordinate system with the origin at the center of mass of the drop. Consequently, boundary conditions (5) and (7) can be expanded into a Taylor series in the vicinity of an equivalent spherical drop of radius R. Then, boundary condition (7) takes the form

$$\Phi(\mathbf{r}, t)|_{r = 1 + f(\theta, t)} \approx \Phi(\mathbf{r}, t)|_{r=1} + (f(\theta, t)\boldsymbol{\nabla})\Phi(\mathbf{r}, t)|_{r=1}$$

$$+ \frac{1}{2}(f(\theta, t)\nabla)^2\Phi(\mathbf{r}, t)|_{r=1} + O(\varepsilon^3) = 0; \qquad (16)$$

$$f(\theta, t) \equiv \frac{1}{3}e^2P_2(\mu) + \xi(\theta, t).$$

(i) For corrections $\Phi^{(1)}$ of order $\sim\varepsilon^{3/2}$ to the potential, the problem is stated as follows:

$$\Delta\Phi^{(1)} = 0;$$

$$r \longrightarrow \infty\colon \Phi^{(1)} \longrightarrow 0;$$

$$r = 1\colon \Phi^{(1)} = -\frac{\partial\Phi^{(0,0)}}{\partial r}\xi^{(1)};$$

$$\int_0^\pi\left(\frac{\partial\Phi^{(1)}}{\partial r}\right|_{r=1}$$

$$\left.+\left(\frac{\partial^2\Phi^{(0,0)}}{\partial r^2} + 2\frac{\partial\Phi^{(0,0)}}{\partial r}\right)\right|_{r=1}\xi^{(1)}\Bigg)\sin(\theta)d\theta = 0.$$

A solution has the form

$$\Phi^{(1)}(r, \theta, t) = \left(\frac{16\pi}{9}\beta\right)^{1/2}\left[3\sum_{n=1}^{\infty}H_n^{(1)}(t)r^{-(n+1)}P_n(\mu)\right]; \qquad (17)$$

$$H_n^{(1)}(t) = (K_{1,n+1,n}M_{n+1}^{(1)}(t) + K_{1,n-1,n}M_{n-1}^{(1)}(t)).$$

(ii) In order of smallness of $\sim\varepsilon^{5/2}$, the electrostatic problem for correction $\Phi^{(2)}$ to the potential is written in the form

$$\Delta\Phi^{(2)} = 0;$$

$$r \longrightarrow \infty\colon \Phi^{(2)} \longrightarrow 0;$$

$$r = 1\colon \Phi^{(2)} = -\frac{\partial\Phi^{(0,0)}}{\partial r}\xi^{(2)} - \left[\frac{\partial^2\Phi^{(0,0)}}{\partial r^2}\left(\frac{\beta}{3}P_2(\mu) + \frac{1}{2}\xi^{(1)}\right)\right.$$

$$\left.+\beta\frac{\partial\Phi^{(0,1)}}{\partial r}\right]\xi^{(1)} - \frac{\partial\Phi^{(1)}}{\partial r}\left(\frac{\beta}{3}P_2(\mu) + \xi^{(1)}\right);$$

$$\int_0^\pi\left[\frac{\partial\Phi^{(2)}}{\partial r} + \left(\frac{\partial^2\Phi^{(0,0)}}{\partial r^2} + 2\frac{\partial\Phi^{(0,0)}}{\partial r}\right)\xi^{(2)}\right.$$

$$+ \beta\left(\frac{\partial^2\Phi^{(0,1)}}{\partial r^2} + 2\frac{\partial\Phi^{0,1}}{\partial r}\right)\xi^{(1)} + \left(\frac{\partial^3\Phi^{(0,0)}}{\partial r^3}\right.$$

$$\left.+ 4\frac{\partial^2\Phi^{(0,0)}}{\partial r^2} + 2\frac{\partial\Phi^{(0,0)}}{\partial r}\right)\left(\frac{\beta}{3}P_2(\mu) + \frac{1}{2}\xi^{(1)}\right)\xi^{(1)}$$

$$+ \left(\frac{\partial^2\Phi^{(1)}}{\partial r^2} + 2\frac{\partial\Phi^{(1)}}{\partial r}\right)\left(\frac{\beta}{3}P_2(\mu) + \xi^{(1)}\right)$$

$$- \frac{\beta}{3}\left(\frac{\partial^2\Phi^{(0,0)}}{\partial r\partial\theta}\xi^{(1)} + \frac{\partial\Phi^{(1)}}{\partial\theta}\right)\frac{\partial P_2(\mu)}{\partial\theta}$$

$$-\left(\frac{\partial^2\Phi^{(0,0)}}{\partial r\partial\theta}\left(\xi^{(1)} + \frac{\beta}{3}P_2(\mu)\right) + \beta\frac{\partial\Phi^{(0,1)}}{\partial\theta}\right.$$

$$\left.\left.+ \frac{\partial\Phi^{(1)}}{\partial\theta}\right)\frac{\partial\xi^{(1)}}{\partial\theta}\right]\Bigg|_{r=1}\sin(\theta)d\theta = 0.$$

A solution to this problem has the form

$$\Phi^{(2)}(r, \theta, t) = \left(\frac{16\pi}{9}\beta\right)^{1/2}\sum_{n=1}^{\infty}H_n^{(2)}(t)r^{-(n+1)}P_n(\mu);$$

$$H_n^{(2)}(t) = 3[K_{1,n+1,n}M_{n+1}^{(2)}(t) + K_{1,n-1,n}M_{n-1}^{(2)}(t)$$

$$+ \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(K_{1,i+1,i}M_{i+1}^{(1)}(t) + K_{1,i-1,i}M_{i-1}^{(1)}(t)$$

$$- (K_{i,j,n+1}K_{1,n+1,n} + K_{i,j,n-1}K_{1,n-1,n})M_i^{(1)}(t))$$

$$\times(i+1)K_{i,j,n}M_j^{(1)}(t)] +$$

$$+ \beta\left[\left(K_{1,n+1,n}\left((n+3)K_{2,n,n} - \frac{4}{5}\right)\right.\right. \qquad (18)$$

$$\left.+ (n+5)K_{1,n+1,n+2}K_{2,n+2,n}\right)M_{n+1}^{(1)}(t) + \left(K_{1,n-1,n}\right.$$

$$\times\left((n+3)K_{2,n,n}-\frac{4}{5}\right)+(n+1)K_{1,n-1,n-2}K_{2,n-2,n}\Bigg)$$

$$\times M_{n-1}^{(1)}(t)+(n+5)K_{1,n+3,n+2}K_{2,n+2,n}M_{n+3}^{(1)}(t)$$

$$+(n+1)K_{1,n-3,n-2}K_{2,n-2,n}M_{n-3}^{(1)}(t)\Bigg].$$

In the above expressions for $\Phi^{(1)}(r, \theta, t)$ and $\Phi^{(2)}(r, \theta, t)$, it is implied that the subscripts at amplitudes $M_n^{(m)}(t)$ ($m = 1, 2$) cannot be negative (the corresponding terms must be rejected).

(iii) A final expression for the electrostatic potential near a nonlinearly vibrating conducting drop in the desired order of smallness takes the form

$$\Phi(r,\theta,t)$$

$$=\frac{E_0}{r^2}\cos(\theta)\left[1-r^3+\frac{e^2}{20r^2}(8r^2-3+15\cos(2\theta))\right] \quad (19)$$

$$+E_0\sum_{n=1}^{\infty}[3\varepsilon H_n^{(1)}(t)+\varepsilon^2H_n^{(2)}(t)]r^{-(n+1)}P_n(\mu).$$

(7) An expression for the electric field strength near the drop is easily found from the relationship $\mathbf{E} = -\mathrm{grad}(\Phi)$,

$$\mathbf{E} = E_0\Bigg\{\left[1+\frac{2}{r^3}+\frac{e^2}{5r^5}(4r^2-3+15\cos(2\theta))\right]\cos(\theta)$$

$$+\varepsilon\sum_{n=1}^{\infty}[3H_n^{(1)}(t)+\varepsilon H_n^{(2)}(t)](n+1)r^{-(n+2)}P_n(\mu)\Bigg\}\mathbf{n}_r$$

$$+E_0\Bigg\{\left(-1+\frac{1}{r^3}+\frac{e^2}{20r^5}(8r^2-3+15\cos(2\theta))\right)\sin(\theta)$$

$$+\frac{3e^2}{2r^5}\sin(2\theta)\cos(\theta)+\sum_{n=1}^{\infty}[3\varepsilon H_n^{(1)}(t)+\varepsilon^2H_n^{(2)}(t)]$$

$$\times(n+1)r^{-(n+2)}-\frac{\partial P_n(\mu)}{\partial\theta}\Bigg\}\mathbf{n}_\theta.$$

Immediately on the surface of a conducting drop, $r(\theta, t) = 1 + f(\theta, t)$, only the field component normal to the surface is other than zero. In the spherical coordinate system, it has the form

$$E_n(\theta,t) = 3E_0\cos(\theta)+3E_0\cos(\theta)\left(e^2\cos\left((\theta)^2-\frac{3}{5}\right)\right.$$

$$\left.-\varepsilon2\xi^{(1)}(\theta,t)\right)-\varepsilon^{3/2}\frac{\partial\Phi^{(1)}}{\partial r}\bigg|_{r=1}$$

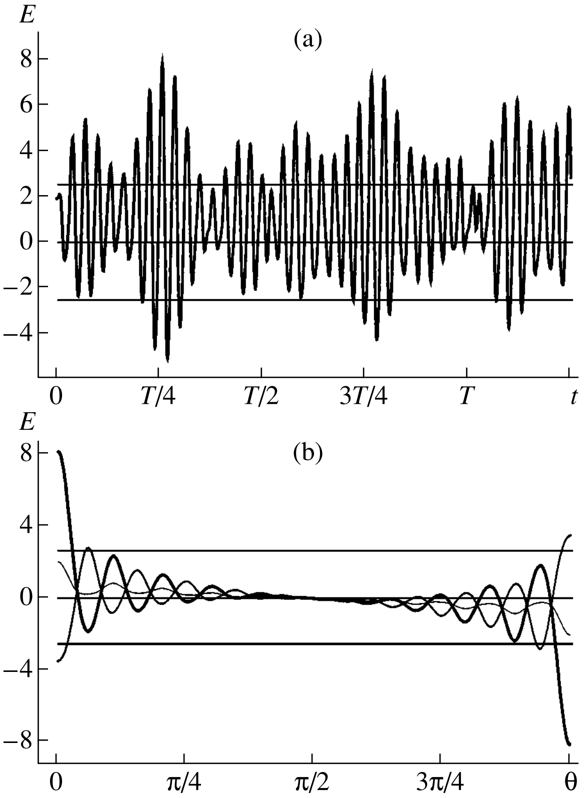


Fig. 1. (a) Dimensionless electric field strength E at one of the tops of the drop vs. dimensionless time for the case when the initial deformation is specified by superposition of the fundamental and 20th modes for $E_0 \approx 0.167$, $h_2 = 0.75$, and $h_{20} = 0.25$. (b) Electric field strength vs. angle θ at various time instants: the thin line, zero time; medium-thickness line, $t = 0.75T$; and thick line, $t = 0.5645T$ (the instant the amplitude reaches a maximum at the axis of symmetry). T is the vibration period of the fundamental mode.

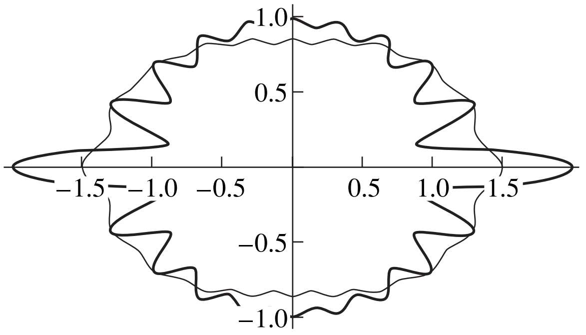


Fig. 2. Generatrix of the drop at the instant the electric field strength reaches a maximum (thick line). The initial shape of the drop (thin line) is specified by superposition of the virtually excited fundamental and 20th modes. The calculations were performed for the same values of the physical parameters as in Fig. 1.

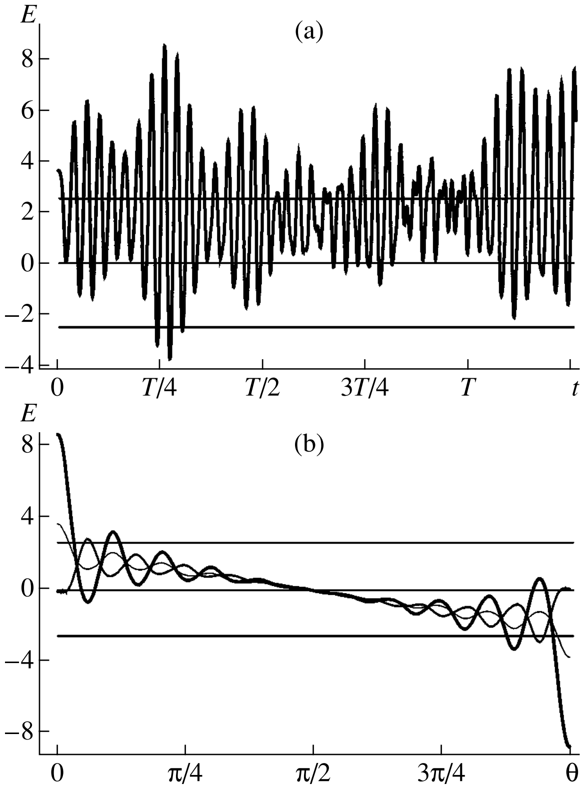


Fig. 3. (a, b) The same dependences as in Fig. 1 calculated for $E_0 \approx 0.53$ and halved vibration amplitude.

$$+E_0\Big[e^2\frac{1}{5}\cos(\theta)\Big(\frac{e^2}{4}(315\cos(\theta)^4-256\cos(\theta)^2+37)+\varepsilon(75\cos(\theta)^2-43)\xi^{(1)}\Big)+\varepsilon^2\frac{3}{2}\Big(\cos(\theta)\times\Big(4\xi^{(2)}-8(\xi^{(1)})^2+\Big(\frac{\partial\xi^{(1)}}{\partial\theta}\Big)^2\Big)-2\sin(\theta)\xi^{(1)}\frac{\partial\xi^{(1)}}{\partial\theta}\Big)\Big] \qquad (20)$$

$$-\varepsilon^{5/2}\Big[\frac{\partial\Phi^{(2)}}{\partial r}+\xi^{(1)}\frac{\partial^2\Phi^{(1)}}{\partial r^2}-\frac{\partial\xi^{(1)}}{\partial\theta}\frac{\partial\Phi^{(1)}}{\partial\theta}\Big]\Big|_{r=1}+\varepsilon^{3/2}\Big[e^2\frac{1}{6}(1-3\cos(\theta)^2)\frac{\partial^2\Phi^{(1)}}{\partial r^2}-6\cos(\theta)\sin(\theta)\frac{\partial\Phi^{(1)}}{\partial\theta}\Big]\Big|_{r=1};$$

$$\xi^{(1)}\equiv\sum_{n=0}^{\infty}M_n^{(1)}(t)P_n(\mu);\qquad \xi^{(2)}\equiv\sum_{n=0}^{\infty}M_n^{(2)}(t)P_n(\mu),$$

where functions $\Phi^{(1)}(r, \theta, t)$ and $\Phi^{(2)}(r, \theta, t)$ are defined by expressions (17) and (18).

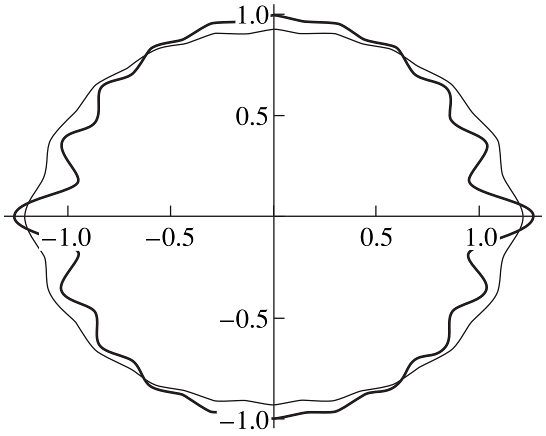


Fig. 4. The same as in Fig 2 for the same values of the physical parameters as in Fig. 3.

When writing (20), we used expansion (16) for the boundary conditions.

(8) Figure 1 shows the results calculated for the electric field strength near the drop using (20). Specifically, Fig. 1a is the time evolution of the field at one of the tops of the drop (at the other top, the field differs only in sign) and Fig. 1b plots the field against angle θ at different time instants. In the calculation, the dimensionless strength of the external uniform electrostatic field was taken to be equal to ≈0.167, which corresponds to dimensional field $E_0 \approx 1.2$ kV/cm for water drops with a radius of 100 μm (such fields are measured in full-scale experiments in storm clouds [4]). Such a value of the electrostatic field strength provides an equilibrium spheroidal deformation of the drop with an eccentricity squared equal to $e^2 \approx 5 \times 10^{-3}$.

It was also assumed that the initial deformation of the drop causing nonlinear vibrations is specified by superposition of the fundamental ($n = 2$) and 20th ($n = 20$) modes. Vibrations at the fundamental mode (spheroidal vibrations) with an amplitude close to the radius of the drop are detected in full-scale measurements [16] and are explained by interaction of the drop with vortices separated from its surface in the turbulent air. Higher modes are excited in large drops (with radius $R \approx 100$ μm) of storm clouds by collisions with fine droplets ($r \approx 10$ μm), whose concentration in clouds is insignificant [4]. In the calculations, the dimensionless vibration amplitude of the fundamental mode was set equal to ≈0.3 and that of the 20th mode was estimated from the ratio of the colliding drop radii as $(r/R) \approx 0.1$. In Fig. 2 shows the generatrix of the axisymmetric drop for the parameter values mentioned above and various time instants.

The straight lines at $E = \pm 2.52$ in Fig. 1 show the dimensionless field strength that is critical for corona discharge initiation. It is easy to see that the field due to the charge induced at the tops of the drop is sufficient for corona initiation. It is worthy to note that the strength of the external uniform electrostatic field used in the calculations is only a tenth of the critical value at which the drop becomes unstable against the induced charge [10, 17].

Figure 3 demonstrates dependences similar to those shown in Fig. 1. Here, external uniform electrostatic field E_0 is $\sqrt{10}$ times higher than the strength used in the calculations shown in Fig. 1, while the vibration amplitude is half as large. Figure 4 shows the drop's generatrix corresponding to the calculations presented in Fig. 3.

CONCLUSIONS

Thus, the electric field strength at the tops of a conducting liquid drop nonlinearly vibrating in external uniform electrostatic field E_0 was calculated. It is shown that, for those values of E_0 observed in full-scale measurements in storm clouds, the field at the top of the drop may exceed the critical corona-initiating value. For this reason, a detailed consideration of lighting discharge initiation by a corona near a large drop seems to be topical.

REFERENCES

1. V. A. D'yachuk and V. M. Muchnik, Dokl. Akad. Nauk SSSR **248**, 60 (1979).
2. A. I. Grigor'ev and S. O. Shiryaeva, Phys. Scr. **54**, 660 (1996).
3. A. I. Grigor'ev, S. O. Shiryaeva, and M. V. Volkova, Zh. Tekh. Fiz. **73** (11), 31 (2003) [Tech. Phys. **48**, 1389 (2003)].
4. I. P. Mazin, A. Kh. Khrgian, and I. M. Imyanitov, *Clouds and Cloudy Atmosphere: A Handbook* (Gidrometeoizdat, Leningrad, 1989) [in Russian].
5. J. A. Tsamopoulos and R. A. Brown, J. Fluid Mech. **147**, 373 (1984).
6. Z. Feng, J. Fluid Mech. **333**, 1 (1997).
7. S. O. Shiryaeva, Zh. Tekh. Fiz. **72** (4), 15 (2002) [Tech. Phys. **47**, 389 (2002)].
8. S. O. Shiryaeva, A. N. Zharov, and A. I. Grigor'ev, Zh. Tekh. Fiz. **74** (1), 10 (2004) [Tech. Phys. **49**, 8 (2004)].
9. K. J. Cheng, Phys. Lett. A **112A**, 392 (1985).
10. A. I. Grigor'ev and O. A. Sinkevich, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 6, 10 (1985).
11. Z. C. Feng, Quarterly Appl. Math. **47**, 555 (1990).
12. Z. C. Feng and K. V. Beard, Proc. R. Soc. London **430**, 133 (1990).
13. Z. C. Feng and K. V. Beard, J. Fluid Mech. **227**, 429 (1991).
14. S. O. Shiryaeva, Zh. Tekh. Fiz. **66** (9), 12 (1996) [Tech. Phys. **41**, 865 (1996)].
15. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1992; Pergamon, New York, 1984).
16. K. V. Beard, Rev. Geophys. **25**, 357 (1987).
17. G. L. Taylor, Proc. R. Soc. London A **280**, 383 (1964).

Translated by N. Mende

Technical Physics, Vol. 50, No. 7, 2005, pp. 864–867. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 48–51.
Original Russian Text Copyright © 2005 by Porubov.

GASES AND LIQUIDS

On the Localization of 2D Nonlinear Internal Waves in a Two-Layer Fluid

A. V. Porubov

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
e-mail: porubov@math.ioffe.ru

Received September 27, 2004

Abstract—A 2D generalized Gardner equation is used to describe 2D nonlinear internal waves in a two-layer fluid. Unlike the previous model based on the Kadomtsev–Petviashvili equation, the model considered here allows for the instability of a plane internal solitary wave. Such a possibility causes the wave to be localized in any direction. Relationships between the thicknesses and densities of the layers under the instability conditions are obtained. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Nonlinear wave processes in a two-layer fluid are of great practical importance, primarily, in oceanology. Of special interest are localized or solitary waves, which usually result from the balance between nonlinearity and dispersion. Having originated from an initial perturbation, they are capable of traveling with invariable shape and velocity. Having a finite, sometimes significant, amplitude, they can transfer a considerable amount of energy over long distances. The localization of the initial perturbation may be attended by a rise in the wave amplitude. Along with localization in the direction of propagation, which gives rise to a plane solitary wave, localization in the transverse direction is also possible. In the latter case, a rise in the amplitude may far exceed that during the plane wave formation.

It is known that solitary waves as solutions to different equations have different properties. The type of model equation for both surface and internal waves considerably depends on the ratios between the thicknesses and densities of fluid layers [1]. In particular, for comparable thicknesses, the long-wave limit is described by the one-dimensional Korteweg de Vries equation [1–5]. However, experiments often cast doubt on the adequacy of this description. To remedy the situation, it was suggested [3–5] that cubic nonlinearity be taken into account. Then, model equations take the form of the modified Korteweg de Vries equation or the Gardner equation with cubic and quadratic nonlinearities.

Two-dimensional localized constant-shape waves can form via interaction of plane waves propagating at an angle to each other or result from the balance between nonlinearity, dispersion, and diffraction. A comprehensive analytical and numerical study of the first opportunity is given, for example, in [6, 7] for the most popular Kadomtsev–Petviashvili (KP) 2D model equation. The second can be described either in terms of transverse instability analysis [8, 9] or by an exact solution in the form of “lump” [1, 10]. In terms of the KP equation, the 2D localization of a wave is accompanied by a considerable rise in its amplitude, as was established analytically for an initial plane solitary wave in [11]. An even higher rise in amplitude is predicted when an initial localized perturbation evolves into a lump in solving the KP equation [12].

As applied to internal waves, the KP equation was derived earlier (see, e.g., [1, 2]). It seems of interest to investigate the transverse instability of a plane solitary internal wave in terms of a model based on the Gardner equation. In this study, we (i) generalize the Gardner equation to 2D internal waves in a two-layer fluid; (ii) study the transverse instability of a solitary wave resulting from the Gardner equation; and (iii) establish a correlation between the instability conditions, thicknesses and densities of the layers, and the sign of the solitary wave amplitude.

2D GENERALIZATION OF THE GARDNER EQUATION

Consider a system of two horizontal layers of immiscible ideal fluids (the lower and upper layers have constant thickness H and H', respectively). For simplicity, the upper boundary of the upper layer is assumed to be unperturbed and the densities of fluids are taken to be constant. The case of the perturbed upper boundary was studied earlier [3], and the types of stratification were considered in [5]. The plane $z = 0$ of the Cartesian system coincides with the unperturbed interface. The constant density of the lower fluid is denoted by ρ and the components of the velocity vector along axes x, y, and z, by $u(x, y, z, t)$, $v(x, y, z, t)$, and $w(x, y, z, t)$, respectively. Similarly, for the upper fluid, we have ρ', $u'(x, y, z, t)$, $v'(x, y, z, t)$, and $w'(x, y, z, t)$, where t is the time.

Let $\eta(x, y, t)$ be the perturbation of the interface. Then, the lower and upper fluids occupy the domains $-H < z < \eta$ and $\eta < z < H'$, respectively.

It seems convenient to introduce velocity potentials $u = \Phi_x$, $v = \Phi_y$, $w = \Phi_z$, $u' = \Phi'_x$, $v' = \Phi'_y$, and $w' = \Phi'_z$ into the equations and boundary conditions. As a result, we have

$$\Phi_{xx} + \Phi_{yy} + \Phi_{zz} = 0 \quad \text{at} \quad -H < z < \eta, \tag{1}$$

$$\Phi'_{xx} + \Phi'_{yy} + \Phi'_{zz} = 0 \quad \text{at} \quad \eta < z < H', \tag{2}$$

$$\Phi_z = 0 \quad \text{at} \quad z = -H, \tag{3}$$

$$\Phi'_z = 0 \quad \text{at} \quad z = H'; \tag{4}$$

while for $z = \eta$,

$$\begin{aligned} &\rho(\Phi_t + 1/2(\Phi_x^2 + \Phi_y^2 + \Phi_z^2) + g\eta) \\ &= \rho'(\Phi'_t + 1/2(\Phi'^2_x + \Phi'^2_y + \Phi'^2_z) + g\eta), \end{aligned} \tag{5}$$

$$\eta_t + \Phi_x\eta_x + \Phi_y\eta_y = \Phi_z, \tag{6}$$

$$\eta_t + \Phi'_x\eta_x + \Phi'_y\eta_y = \Phi'_z. \tag{7}$$

To derive a model equation, we impose a number of restrictions. Only long waves with typical length L and small, while finite, amplitude A are considered. Accordingly, we introduce the following scales for the variables: L for x, Y for y, H for z, $L/\sqrt{gH}$ for t, A for η, and $AL\sqrt{gH}/H$ for Φ and Φ'. We denote $\sigma = H'/H$ and $\Delta = \rho'/\rho$, introduce small parameter $\varepsilon = H/L$, and assume that variations in the y direction are weaker than in the x direction ($Y = L/\varepsilon$). Then, the solutions to dimensionless Laplace equations (1) and (2) with boundary conditions (3) and (4) have the form

$$\begin{aligned} \Phi &= \phi(x, y, t) - \varepsilon^2\frac{(z+1)^2}{2}\phi_{xx} \\ &+ \varepsilon^4\left(\frac{(z+1)^4}{24}\phi_{xxxx} - \frac{(z+1)^2}{2}\phi_{yy}\right) + O(\varepsilon^6), \end{aligned} \tag{8}$$

$$\begin{aligned} \Phi' &= \phi'(x, y, t) - \varepsilon^2\frac{(z-\sigma)^2}{2}\phi'_{xx} \\ &+ \varepsilon^4\left(\frac{(z-\sigma)^4}{24}\phi'_{xxxx} - \frac{(z-\sigma)^2}{2}\phi'_{yy}\right) + O(\varepsilon^6), \end{aligned} \tag{9}$$

where $\phi(x, y, t)$ and $\phi'(x, y, t)$ are new unknown functions.

Consider first the case $A/H = O(\varepsilon^2)$. To derive a model equation, we assume that a solution depends on phase variable $\theta = x - vt$, y, and slow time $\tau = \varepsilon^2 t$. Substituting Eqs. (8) and (9) into dimensionless boundary conditions (5)–(7), we find that phase velocity v is given by

$$v^2 = \frac{\sigma(1-\Delta)}{\sigma+\Delta}, \tag{10}$$

and perturbation η of the interface is described by the KP equation [8]

$$(\eta_\tau + a^*\eta_\theta^2 + b\eta_{\theta\theta\theta})_\theta + d\eta_{yy} = 0, \tag{11}$$

where

$$a^* = \frac{3v(\sigma^2-\Delta)}{4\sigma(\sigma+\Delta)}, \quad b = \frac{\sigma v(1+\sigma\Delta)}{6(\sigma+\Delta)}, \quad d = \frac{v}{2}.$$

It is easy to see that $b > 0$ and $d > 0$ and that the coefficient multiplying the square nonlinearity may be of either sign. In particular, if the ratio between the thicknesses of the layers is $\sigma = \sigma^*$,

$$\sigma^* = \sqrt{\Delta}, \tag{12}$$

we have $a^* = 0$. In this case, the nonlinearity–dispersion balance, which is the condition for wave localization, is impossible for the relationship between A/H and ε accepted. Therefore, we change this relationship in the initial dimensionless equations to $A/H = O(\varepsilon)$ and also put $v = v^* + \varepsilon v_1 + \ldots$, and $\sigma = \sigma^* + \varepsilon\sigma_1 + \ldots$. Solving the problem in the zeroth approximation leads to Eq. (10) for v^*, which, in view of (12), takes the form

$$v^{*2} = 1 - \sigma^*. \tag{13}$$

The first-order correction to the velocity has the form

$$v_1 = \frac{\sigma_1 v^*}{2(\sigma^*+1)}.$$

The square and cubic nonlinearities turn out to be of the same order of smallness; then, from the condition of solvability of the next-order approximation, we arrive at a 2D generalization of the Gardner equation for perturbation η of the fluid surface in the form

$$(\eta_\tau + a\eta_\theta^2 + c\eta_\theta^3 + b\eta_{\theta\theta\theta})_\theta + d\eta_{yy} = 0, \tag{14}$$

where

$$a = \frac{3v^*\sigma_1}{2\sigma^*(\sigma^*+1)}, \quad b = \frac{v^*(1+\sigma^*+\sigma^{*2})}{6\sigma^*},$$

$$c = -\frac{3v^*}{\sigma^*}, \quad d = \frac{v^*}{2}.$$

We notice that $b > 0$ and $d > 0$ again, the coefficient multiplying the cubic nonlinearity is always negative, $c < 0$, and the sign of a is defined by the sign of σ_1.

STABILITY OF AN INTERNAL PLANE WAVE

When $b > 0$ and $d > 0$, the KP equation has a stable solution in the form of a plane solitary wave, while a

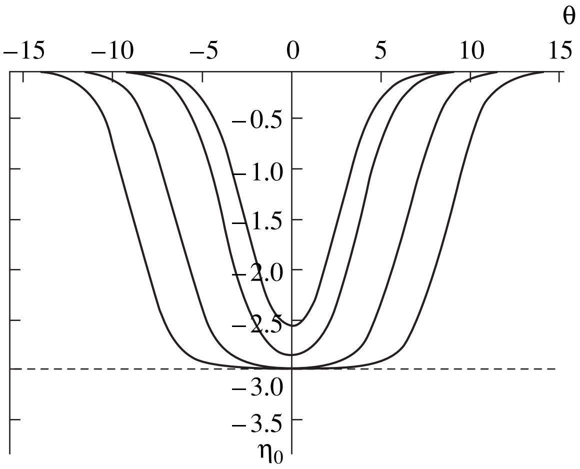


Fig. 1. Evolution of the Gardner solitary wave at $k \longrightarrow \sqrt{-2a^2/(9bc)}$ in the projection onto the plane $y = 0$. The dashed line shows the ultimate amplitude.

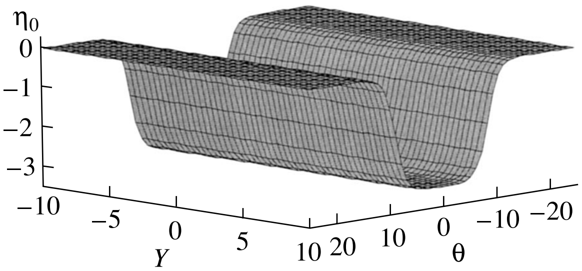


Fig. 2. Plane solitary trough following from the Gardner equation.

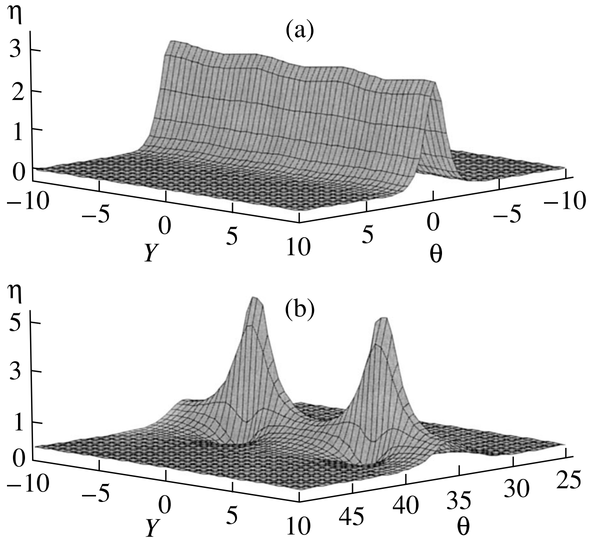


Fig. 3. Transformation of (a) the Kadomtsev–Petviashvili plane solitary wave into (b) a sequence of 2D localized waves.

solution in the form of a 2D localized lumped wave can take place at $b > 0$, $d < 0$ or $b < 0$, $d > 0$ [1, 8].

At $a = 0$ ($\sigma_1 = 0$), Eq. (14) transforms, in the one-dimensional approximation, to the modified Korteweg de Vries equation for function η. For the adopted signs of the nonlinear cubic and linear dispersion terms, this equation does not have a solution in the form of a localized wave vanishing at infinity [1–3].

The well-known single-soliton solution to the Gardner equation has the form [3, 13]

$$\eta_0 = \frac{3bk^2}{a(B_1\cosh(k\xi)+1)}, \tag{15}$$

where

$$B_1 = \sqrt{1+\frac{9bck^2}{2a^2}}, \quad \xi = \theta - bk^2\tau.$$

At $k \longrightarrow \sqrt{-2a^2/(9bc)}$, solution (15) tends to acquire the shape of an extended trough; accordingly, the wave amplitude tends toward the limiting value $-2a/3c$ (Fig. 1). This mode is evidently realized in the case at hand $bc < 0$. In the 2D statement, Eq. (15) describes the propagation of a plane trough (Fig. 2). It is important to know whether such a wave is stable against transverse perturbations. Let a solution to Eq. (14) have the form

$$\eta(\theta, \tau, y) = \eta_0 + \delta w(\theta, \tau)Exp(\lambda\tau + ipy),$$

where $\delta \ll 1$.

Then, correction w is found from the linear equation

$$\begin{gathered}(w_\tau + 2a(\eta_0 w)_\theta + 3c(\eta_0^2 w)_\theta + bw_{\theta\theta\theta})_\theta \\ + \lambda w_\theta - p^2 dw = 0,\end{gathered} \tag{16}$$

whose asymptotic solution in the case of small p (weak transverse modulation) is sought in the form

$$w = w_0 + pw_1 + p^2 w_2 + \ldots, \quad \lambda = p\lambda_1 + p^2\lambda_2 + \ldots.$$

Substituting these power series into Eq. (16) and equating terms with the same powers of p, we obtain

$$w_0 = \eta_{0,\theta},$$

$$w_1 = \frac{3\lambda_1(B_1^2 k\xi\sinh(k\xi) - (1+B_1^2)\cosh(k\xi) - 2B_1)}{2aB_1(1+B_1\cosh(k\xi))^2}.$$

From the condition of solvability in the p^2 approximation, an equation for λ_1 follows,

$$\lambda_1^2 = \frac{a^2 dB_1^2}{2c(1-B_1^2)}F,$$

where

$$F = 1 - \frac{2}{\sqrt{1-B_1^2}} \arctan\sqrt{(1-B_1)/(1+B_1)}.$$

It is easy to check that $F < 0$ for allowable values of B_1 $(0 < B_1 < 1)$. Hence, whether the wave is stable or unstable depends on the sign of c: it is unstable for negative c and stable for positive c.

CONCLUSIONS

Thus, we generalized the Gardner equation for internal waves in a two-layer fluid to the 2D case and derived an explicit dependence of the transverse instability of a plane solitary wave on the coefficients involved in the equation. The instability thus obtained depends on coefficient a multiplying the nonlinear cubic term, as distinct from the instability following from the Kadomtsev–Petviashvili equation. The sign of coefficient a depends on the sign of σ_1, i.e., on the ratio between the thicknesses and densities of the layers, as follows from the definition of σ and σ^*,

$$\varepsilon\sigma_1 = \frac{H'}{H} - \sqrt{\frac{\rho'}{\rho}}.$$

In addition, the sign of a defines that of the amplitude of solitary wave (15); namely, when $a < 0$, only waves with troughs are unstable. Taking into account surface tension at the interface may change the sign of coefficient b in Kadomtsev–Petviashvili equation (11) to negative. Then, the condition $a^* > 0$ corresponds to unstable troughs and $a^* < 0$, to unstable hump-shaped solitons.

In the unstable case, one may expect the formation of 2D localized troughs similarly to the evolution of a plane hump resulting from Kadomtsev–Petviashvili equation (11) at $b < 0$ and $a^* < 0$ [11]. Figure 3 presents the initial and final phases of this solution. Importantly, the wave amplitude grows during this process. As regards to the formation of the plane wave itself, the formation of solitary wave (15) resulting from the one-dimensional Gardner equation was numerically analyzed for an arbitrary initial perturbation in [13]. It turned out that, unlike ordinary solitons, wave (15) may appear only if the initial pulse (perturbation) is sufficiently wide. Moreover, only one wave may arise in this case rather than a sequence of solitons with various amplitudes, as follows, for example, from the Korteweg de Vries equation.

Note finally that we restricted our consideration to the unperturbed free surface of the upper layer and constant densities. Both the perturbation of the surface and stratification can be taken into consideration by generalizing the results of [3, 5] for the 2D case.

REFERENCES

1. M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981; Mir, Moscow, 1987).
2. V. D. Djordjevic and L. G. Redekopp, J. Phys. Oceanogr. **8**, 1016 (1978).
3. T. Kakutani and N. Yamasaki, J. Phys. Soc. Jpn. **45**, 674 (1978).
4. J. Miles, Tellus **3.1**, 456 (1979).
5. R. Grimshaw, E. Pelinovsky, and T. Talipova, Nonlinear Proc. Geophys. **4**, 237 (1997).
6. P. Peterson *et al.*, Nonlinear Proc. Geophys. **10**, 1 (2003).
7. H. Tsuji and M. Oikawa, J. Phys. Soc. Jpn. **62**, 3881 (1993).
8. B. B. Kadomtsev and V. I. Petviashvili, Dokl. Akad. Nauk SSSR **192**, 753 (1970) [Sov. Phys. Dokl. **15**, 539 (1970)].
9. J. C. Alexander, R. L. Pego, and R. L. Sachs, Phys. Lett. A **226**, 187 (1997).
10. J. Satsuma and M. J. Ablowitz, J. Math. Phys. **20**, 1496 (1979).
11. K. A. Gorshkov, D. E. Pelinovskii, and Yu. A. Stepanyants, Zh. Éksp. Teor. Fiz. **104**, 387 (1993) [JETP **77**, 237 (1993)].
12. A. A. Minzoni and N. F. Smith, Wave Motion **24**, 291 (1996).
13. A. V. Slyunyaev and E. N. Pelinovskii, Zh. Éksp. Teor. Fiz. **116**, 318 (1999) [JETP **89**, 173 (1999)].

Translated by A. Sidorova

Technical Physics, Vol. 50, No. 7, 2005, pp. 868–875. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 52–59.
Original Russian Text Copyright © 2005 by Temnikov, Orlov, Bolotov, Tkach.

Studies of the Parameters of a Spark Discharge between an Artificial Charged Water-Aerosol Cloud and the Ground

A. G. Temnikov*, A. V. Orlov*, V. N. Bolotov, and Yu. V. Tkach****

** Moscow Power Engineering Institute (Technical University),*
ul. Krasnokazarmennaya 17, Moscow, 111250 Russia
e-mail: temnikov@fee.mpei.ac.ru
*** Institute for Electromagnetic Research, Kharkov, 61022 Ukraine*
e-mail: renic@iemr.com.ua

Received October 11, 2004

Abstract—The optical and current characteristics of spark discharges between a grounded electrode and a strongly charged artificial water-aerosol cloud are studied experimentally. The spectral characteristics of the discharge current are investigated using wavelet and Fourier analyses. Three main types of discharge with different final stages are revealed and investigated. It is found that the parameters of a discharge in its final stage depend substantially on the discharge trajectory and the depth to which it penetrates into the aerosol cloud. It is shown that the parameters of the most powerful type of discharge (the brightness of the discharge channel, the current growth rate in the final stage, and the penetration depth into the charged aerosol cloud) are close to the discharge parameters in the main stage of natural lightning. It is also shown that such a discharge neutralizes up to 5% of the cloud charge. In contrast to Fourier analysis, wavelet analysis shows that the signal amplitude (i.e., the energy deposition rate in the discharge channel) is maximum at high frequencies for all types of discharge. Of special interest is the most powerful type of discharge, in which the signal amplitude (which is one to two orders of magnitude larger than in other discharges) is maximum at frequencies of several hundred megahertz. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Studies of discharges in charged water-aerosol clouds are important for both understanding the physics of thunderstorms and predicting the action of lightning on ground and flying objects [1–3]. At present, it is not quite clear what discharge processes occur in different thunderstorm regions and how lightning is initiated and propagates inside the thundercloud and near it. The reason is that the mechanisms for the development of a discharge in strongly charged aerosol clouds (thunderstorm cells) and for the neutralization of the thundercloud charge both during the development of a downward leader and in the main stage of lightning have been examined very poorly (especially experimentally). Experimental facilities [4, 5] do not allow one to produce and study discharges of a length of more than 10 cm.

At present, mechanisms for the interaction between channeled atmospheric discharges and local regions of a charged aerosol cloud (thunderstorm cells) still remain unclear. At the same time, these processes play an important role in the development of lightning and in damaging ground and flying objects by it [1, 3]. It is commonly accepted that the first component of an upward lightning proceeds without the main stage, which takes place only in the subsequent components that start from the cloud, propagate toward the ground, and are essentially identical to the components of downward lightning [1]. However, not every upward lightning penetrating into a thundercloud is accompanied by a return stroke [3, 6, 7]; i.e., the start-up of the "dart leader–return stroke" system is determined by the interaction between an upward leader and a thunderstorm cell. Moreover, an important role in the development of different types of bipolar lightning is played by upward discharges and the processes accompanying the development of the discharge inside the cloud, as well as by the switching of different regions of the charged cloud to the lightning discharge [8].

Another problem is related to determining mechanisms for the development of the main stage of lightning, which manifests itself through acoustic and electromagnetic emission [1]. Lightning discharges generate intense radio signals—the so-called "spherics." The frequency spectrum of the spherics spans from several hertz to several hundred megahertz [3, 9]. The frequency spectra of the current in different stages of lightning also lie in this range [10, 11]. Note that the frequency ranges of different types of natural lightning differ substantially from one another. The frequency spectrum of an ordinary cloud-to-ground lightning usually lies in the range 3–30 MHz [12]. The typical frequencies of intracloud lightnings are one order of magnitude higher [13, 14]. The frequency characteristics of discharges between the vertex of a cloud and the ionosphere depend substantially on the type of discharge [15, 16]. The majority of the spectral characteristics of natural lightning were obtained by analyzing the electromagnetic radiation from lightning discharges. As for the spectral characteristics of the current in different stages of various types of lightning, they have been studied in less detail. Basically, there are only data for return strokes and trigger lightning. This is related to difficulties in measuring the lightning current, especially in the intracloud parts of the discharge.

Estimates of the current characteristics of lightning are especially important in analyzing the situation in which an aircraft approaching a thundercloud initiates a bidirectional leader [17]. The parameters of this leader and the subsequent action of lightning on the aircraft are determined by the interaction of one of the leaders with the space charge of the thunderstorm cell that initiates it [18].

The study of strongly charged aerosol clouds capable of initiating channeled discharges allows one to determine the current characteristics of discharges propagating both inside the charged aerosol cloud and between the cloud and ground.

In this paper, we describe and analyze results from experimental studies of the development of a discharge between a negatively charged artificial water-aerosol cloud and a metal rod installed on a grounded plane under the cloud. By varying the parameters of the cloud, one can investigate all the discharge phenomena that can occur during a natural thunderstorm, including the leader and main stages of lightning [19]. Attention is mainly focused on the study of the development of a discharge within a charged water-aerosol cloud. The results from experimental studies of the development of a discharge between a strongly charged artificial water-aerosol cloud and the grounded electrode are used to analyze the amplitude, temporal, and spectral characteristics of the discharge current in different stages of the discharge.

1. EXPERIMENTAL STUDY OF THE OPTICAL AND CURRENT CHARACTERISTICS OF A DISCHARGE BETWEEN A CHARGED WATER-AEROSOL CLOUD AND THE GROUND

The experimental setup consists of a 250-m^3 aerosol chamber, a condensation-type generator of charged aerosol (GCA), an electrode system, and a diagnostic complex. A schematic of the experimental setup is shown in Fig. 1. The GCA is used to produce a charged water-aerosol cloud above a grounded plane. The generator has a converging subsonic nozzle with an outlet diameter of 6 mm. The nozzle unit is enclosed by an electrostatic screen and is installed at a height of 0.8 m above the grounded plane. The axis of the charged aerosol jet can be inclined at an angle of 0°–30° relative to the horizontal grounded plane. These experiments were performed at an inclination angle of 12°. The length of the charged aerosol jet was varied using a large-mesh electrostatic screen oriented perpendicular to the gasdynamic flow. At a pressure in the stream generator of 3–8 atm, the vapor outflow velocity was 400–430 m/s.

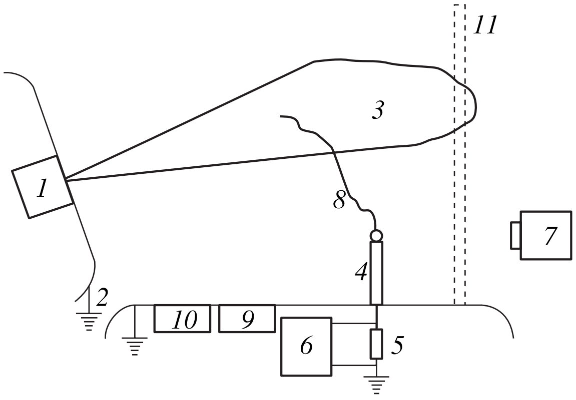


Fig. 1. Schematic of the experimental setup: (*1*) condensation-type generator of charged aerosol, (*2*) grounded electrostatic screens, (*3*) charged aerosol cloud, (*4*) rod electrode with a spherical tip, (*5*) low-inductance shunt, (*6*) Tektronix TDS 754D and TDS 3052 digital oscilloscopes, (*7*) Canon PowerShot G1 digital camera, (*8*) spark discharge, (*9*) electric strength detector, (*10*) dynamic antenna, and (*11*) large-mesh grounded screen.

The volume of the aerosol cloud was a few cubic meters. The charging unit of the GCA produced an output current of up to 150 μA. The space charge density on the axis of the cloud was in the range 1.5×10^{-4}–1.0×10^{-2} C/m^3. The electric potential of the cloud reached 1.5 MV. As a result, a strong electric field was produced between the charged aerosol cloud and the grounded plane. The electric field strength reached 12 kV/cm near the grounded plane and 22 kV/cm at the boundary of the charged aerosol cloud. In the standard operating regime of the experimental setup (i.e., at an output GCA current of 130 μA and a vapor pressure in the steam generator of 5–6 atm), the region of the maximum electric potential (~1.4 MV) lies at a distance of 1.6–1.8 m from the nozzle and at a height of about 1 m above the horizontal grounded plane. At the boundary of the charged aerosol flow, the potential reached 1.0 MV in the cross section that lay at a distance of 1.4–1.5 m from the nozzle. The parameters of the experimental setup and the aerosol clouds produced are described in more detail in [20, 21].

Between the lower boundary of the charged aerosol cloud and the grounded plane, at heights of up to 0.4 m above the horizontal grounded plane and at distances of 1.0–2.0 m from the nozzle, the electric field is almost uniform. It increases rather slowly (by no more than 30–40%) with height: from ~10 kV/cm near the grounded plane to 13–14 kV/cm at a height of 0.4 m. Thus, rod electrodes (or other types of electrodes) of

Table 1. Parameters of different types of discharge between a charged aerosol cloud and a grounded electrode

Parameter	Type I	Type II	Type III	Incomplete discharge
L, m	Up to 2.0	1.0–1.7	1.0–1.3	0.4–0.7
N	2–4	1–2	1	1–2
M	2–4	1–3	1–2	1–2
$I_{\max}$, A	Up to 8	Up to 20	>20	<4
T_f, μs	2–5	0.7–1.9	0.3–1.1	1–4
τ_f, ns (over all the spikes)	50–70	30–50	5–30	100–180
a, kA/μs	0.03–0.09	0.13–0.19	0.4–2.8	~0.02
$Q_{\text{n.f}}$, μC	4.5–6.0	4.0–5.1	5.0–6.5	2.0–3.0
$Q_{\text{n.cl}}$, μC	~2	~1	~2.5–3.0	~0

height less than 0.4 m are exposed to approximately the same electric field.

To initiate a controlled discharge between the aerosol cloud and the ground, a rod ending with a sphere was installed on the plane. The sphere was insulated from the rod and was grounded through a low-inductance shunt. The height of the rod electrode was varied from 100 to 580 mm, and the radius of the sphere was varied from 0.5 to 25 mm. The rod was installed at a distance of 1.3–1.8 m from the nozzle, where the electric field produced by the charged aerosol cloud was the strongest. In the region where the sphere was located, the electric field was almost uniform.

The discharge current signal taken from the low-inductance shunt was fed to a Tektronix TDS 754D and a TDS 3052 digital oscilloscope (Fig. 1). The electric field strength and the charge of the aerosol cloud were measured by a string field meter installed in the grounded plane. Variations in the cloud charge in the course of discharge were measured by a dynamic antenna, the signal from which was fed to the Tektronix TDS 754D oscilloscope. The process of propagation of a discharge and its penetration into the cloud was photographed by a Canon PowerShot G1 digital camera.

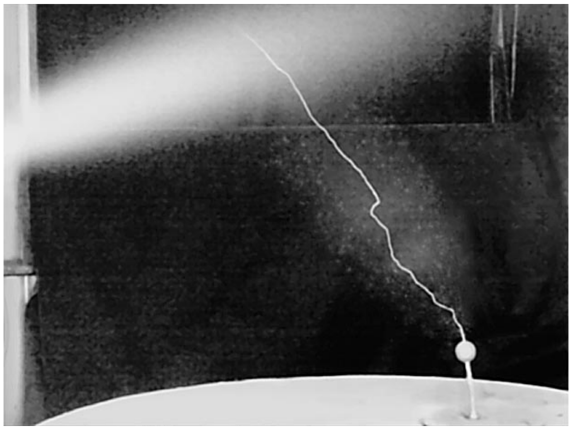


Fig. 2. Final stage of type I discharge.

Three types of spark discharge penetrating into a charged water-aerosol cloud were revealed. The first and third types differed significantly in their optical characteristics and the parameters of the current pulse in the final stage of the discharge, when the space charge of the aerosol cloud was neutralized. Besides these three types, we also observed an incomplete discharge, which did not reach the cloud and was characterized by a feeble glow and small current in its final stage. The optical and current characteristics of different types of discharge (the length L of the spark channel, the number of large segments N, the number of the current peaks M, the maximum discharge current $I_{\max}$, the duration of the final stage T_f, the current rise time τ_f at a level of $(0.3–0.9)I_{\max}$, the current growth rate at a level of $(0.3–0.9)I_{\max}$, the charge $Q_{\text{n.f.}}$ neutralized during the final stage, and the charge of the aerosol cloud $Q_{\text{n.cl.}}$ neutralized by the spark discharge) obtained by processing the data from several hundred discharges are presented in Table 1.

The first type of spark discharge (type I) is characterized by sparks of length up to 2 m, which are feebly glowing even during the final stage of the discharge and consist of several segments. A characteristic feature of type I discharges is the presence of a pronounced bridge at a height of 30–50 cm above the grounded electrode. The bridge is perpendicular to the discharge segments that are connected by it. The length of the bridge is 5–15 cm. An example of type I discharge is shown in Fig. 2. Another important feature of this type of discharge is that it develops only along the periphery of the charged aerosol cloud. The discharge can bypass regions with an elevated space charge density along two paths: along the cloud toward the nozzle screen and along a shorter trajectory that is perpendicular to the jet and passes around the region with the maximal potential. Type I discharges never penetrate into the cloud, but rather propagate along the periphery of the aerosol cloud, where the space charge density is low and the electric field strength is high.

The current waveforms of type I discharges have a multipeak structure (two peaks or more) with an amplitude of up to 8 A. The duration of the final stage is 2–5 μs, and the minimum current rise time is 50–70 ns. The charge neutralized in the final stage of this type of discharge is usually 4.5–6.0 μC. The fraction of the charge neutralized just in the charged aerosol cloud is 30–40% (~2 μC). This is related to the fact that the discharge covers a fairly large distance in the cloud (although along its periphery).

In the final stage of type III discharge, very bright sparks merge into one channel of length 1.0–1.3 m. The final stage of this type of discharge is illustrated in Fig. 3. A specific feature of such a powerful discharge is that it enters the charged aerosol cloud at an angle of 40°–60° with respect to its boundary and penetrates into the cloud to a depth of more than 20–30 cm.

The current waveforms in the final stage of type III discharges have only one peak with an amplitude of more than 20 A. The duration of the final stage of such a discharge is 0.3–1.1 μs, and the minimum current rise time is 5–30 ns. This type of discharge is characterized by higher current growth rate in the final stage: up to ~3 kA/μs, which is by more than one order of magnitude higher than in other types of discharge. The charge neutralized in the final stage of type III discharge is usually 5.0–6.5 μC, and the charge neutralized just in the cloud is ~2.5–3.0 μC. Such a development of the discharge near the boundary of the cloud and inside it can be regarded as "optimal" from the standpoint of maximum neutralization of the charged water-aerosol cloud.

The optical and current characteristics of the final stage of type II discharge lie between those of type I and III discharges. The charge neutralized by this type of discharge is as low as ~1 μC. The reason is that the discharge propagates inside the cloud toward the region with the maximum potential and the electric field strength decreases along the discharge trajectory, while the space charge density of aerosol particles increases. As a result, the discharge propagates inside the cloud in a rapidly decelerating mode. The discharge enters the cloud almost perpendicularly to its boundary, covers a distance of no more than 5–10 cm, and then stops.

2. SPECTRAL CHARACTERISTICS OF THE DISCHARGE CURRENT

To determine the spectral characteristics of the discharge current and its dynamics in the course of discharge, we used Fourier and wavelet analyses.

Wavelet analysis has been successfully used in studying complex unsteady signals. Fourier analysis allows one to obtain the spectrum of the entire signal or certain segments of the signal within given time intervals and thus is appropriate only for steady-state signals.

To trace the evolution of an unsteady signal and obtain the time–frequency representation of the signal, which allows one to analyze the temporal behavior of the signal spectrum, we used not only the window Fourier transformation, but also the wavelet transformation, which possesses a number of important advantages. Wavelet analysis allows one to examine the inner structure of nonuniform objects, such as multiscale or unsteady processes, and to investigate its local properties. Such an approach is especially helpful when studying rapidly varying signals that have both high- and low-frequency components. These capabilities of wavelet analysis have been successfully used to process random signals [22].

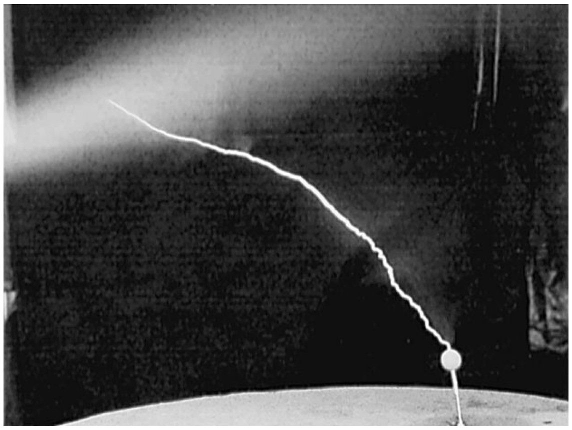


Fig. 3. Final stage of type III discharge.

It should be noted that there are many different types of wavelets; this allows one to choose a wavelet that is most appropriate for studying some particular problem. This is why, together with the wavelet spectrum of the signal, it is always indicated with which type of wavelet it is obtained. The following wavelets are used most frequently: (a) DOG wavelets (derivatives of a Gaussian function), (b) COS and SIN Morlet wavelets, (c) complex Morlet wavelets, and (d) Battle–Lemarie wavelets.

In recent years, wavelet analysis has been used to study gas discharges and atmospheric electricity. In studying thunderstorms, wavelet analysis was applied, e.g., to processing signals characterizing variations in the pressure and temperature of air and the parameters of the wind in a thunderstorm cloud. The coefficients of the wavelet transform were then used to describe the thunderstorm activity [23]. A similar approach was employed in [24, 25] to determine the fine structure of the developing clouds.

Wavelet analysis is even more often applied in studying electromagnetic fields generated by lightning [26, 27]. Attempts have been made to identify the elements of the obtained time–frequency spectrum with different stages of lightning: prebreakdown pulses, the return shock, and the main stage of lightning. For this purpose, Haar wavelets [28] are frequently used. These wavelets are successfully employed to analyze discrete signals as well as to study partial discharges in insulating systems [29].

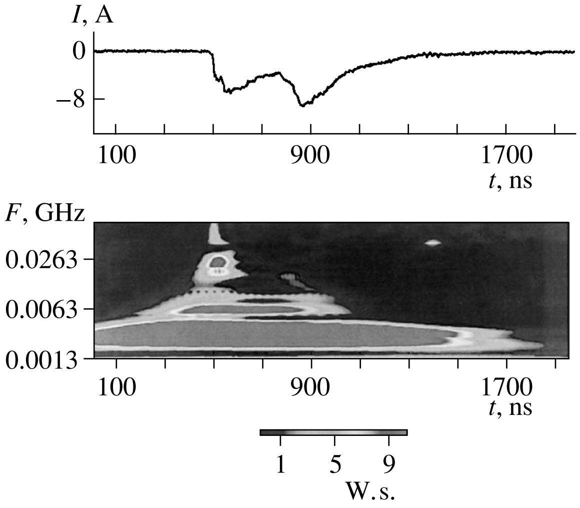


Fig. 4. Wavelet spectrum of type I discharge.

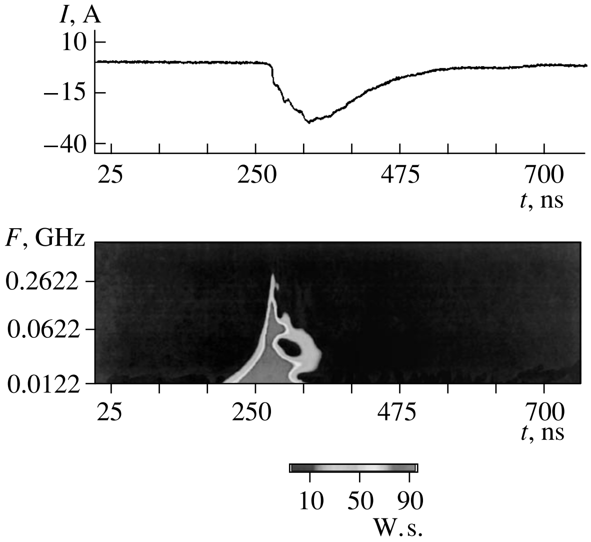


Fig. 5. Wavelet spectrum of type III discharge.

In the present study, the wavelet spectra of the current of a discharge excited between a charged aerosol cloud and the ground are obtained using a specially designed computer program based on complex Morlet wavelets:

$$\psi(t) = \sqrt{\frac{1}{\sqrt{\pi}}}e^{i\frac{2\pi t}{T}}e^{-\frac{t^2}{2T^2}}.$$

The Morlet wavelet contains the parameter T (the oscillation period), which determines the maximum frequency that can be detected using this wavelet.

Such a wavelet transformation can be interpreted as a continuous set of window Fourier transformations with windows of different widths that vary with frequency. This allows one to optimally localize, in frequency and time, all the processes occurring on different time scales. This property is the main advantage of the wavelet transformation as applied to the analysis of multiscale signals, as well as signals undergoing jumps in the amplitude or phase or signals containing several unsteady components with different carrier frequencies (multimode signals). Therefore, this wavelet transformation can be successfully used to analyze the behavior of the discharge current in a charged aerosol cloud.

Figures 4 and 5 present examples of the discharge current waveforms and the relevant wavelet spectra of type I and III discharges. These wavelet spectra correspond to the discharges shown in Figs. 2 and 3. Figure 6 shows the corresponding Fourier spectrum of the type III discharge.

Wavelet analysis of the current waveforms in different types of discharge yielded the characteristic frequencies and the corresponding signal intensities for each type of discharge. Seven characteristic frequency ranges were revealed: <1, 2–5, 6–10, 11–18, 19–30, 31–50, 51–150, and 151–300 MHz. The frequency ranges and the corresponding signal intensities (i.e., the rates of energy deposition in a discharge) were found to differ substantially for different types of discharge and for the first and subsequent current pulses in each type of discharge. The results of processing wavelet spectra of the current waveforms in the final stage of discharge, in an incomplete discharge, and in the preceding leader discharge are presented in Table 2.

In the final stage of type I discharges, three frequency ranges (<1, 2–5, and 6–10 MHz) can be distinguished in the wavelet transform of the current signal. These frequency ranges are present throughout the entire final stage, and the signal intensity is maximum in the first frequency range. It should be noted that there is also a more feeble (but rather pronounced) signal at the front of the first pulse in the final stage of the discharge in the frequency ranges of 11–18 and 19–30 MHz.

For an incomplete discharge with a feeble final stage, only two first frequency ranges (<1 and 2–5 MHz) are observed in the current waveform and the signal intensity is a factor of three to five lower than that for the first type of discharge.

Wavelet analysis of the current waveforms shows that, for type II discharges, the frequency range is wider and the signal intensity is higher than those for type I discharges, especially in the frequency ranges of 11–18, 19–30, and 31–50 MHz, which are untypical of type I discharges. Note that the signal in the last frequency range is present only during the first current pulse in the final stage of discharge.

Wavelet analysis shows that the final stage of type III discharge is distinguished among other types of discharge by a wider frequency range and a higher signal intensity at each frequency. Intense signals appear in

Table 2. Results from wavelet analysis for different types of discharge

Characteristic frequencies, MHz	<1	2–5	6–10	11–18	19–30	31–50	51–150	151–300
Type I	+	+	+	+	+	–	–	–
Type II	+	+	+	+	+	+	–	–
Type III	+	+	+	+	+	+	+	+
Incomplete discharge	+	+	–	–	–	–	–	–
Leader discharge	–	–	–	+	+	+	+	+

Table 3. Results from Fourier analysis for different types of discharge

Type of discharge	Incomplete discharge	Type I	Type II	Type III
Characteristic frequencies, MHz	Up to 3–5	Up to 15–25	Up to 25–50	Up to 150–200
Interval between characteristic frequencies in the Fourier spectrum, MHz	0.4–0.6	0.8–2.5	1.8–4.0	5.0–20.0
Signal intensity at a frequency of 5 MHz	10^{-9}	7×10^{-8}	5×10^{-7}	4×10^{-5}

the frequency ranges of 51–150 and 151–300 MHz, which are typical of the leader stage. However, in the final stage of type III discharges, the signal intensity at these frequencies is several hundred times higher than that in the leader. In nature, such intense signals at frequencies of several hundred megahertz were observed during compact (but very powerful) intracloud discharges [14]. In [14], such a strong emission was attributed to the development of powerful intracloud streamer discharges.

It was found that the maximum current growth rate a in the final stage of any type of discharge is nearly proportional to the maximum characteristic frequency $F_{\max}$ in the wavelet spectrum of the discharge current (Fig. 7): the higher the current growth rate, the higher the maximum frequency. It should be noted that such a linear dependence is typical of first current pulses in the final stage of discharge; the current growth rate is maximum just in these pulses in any type of discharge. It was also found that the higher the characteristic frequency of the wavelet spectrum, the higher the signal intensity at this frequency. This result, which was obtained by processing experimental curves using wavelet analysis, differs substantially from the results of Fourier analysis of lightning discharges: generally, the amplitude of the Fourier spectrum rapidly decreases with increasing characteristic frequency [30]. Such a decrease in the amplitude of the Fourier spectrum of the measured signal with increasing characteristic frequency takes place even for very powerful lightning discharges [31]. The same result was obtained for discharges in strongly charged artificial aerosol clouds.

Fourier analysis of the current waveforms also shows that the frequency spectra are very different in different types of discharge. It was found that the more powerful the final stage of a discharge, the larger the interval between the characteristic frequencies in the Fourier spectrum for each type of discharge. The results of Fourier analysis are presented in Table 3.

Along with the intervals between the characteristic frequencies, the frequency ranges in which these frequencies were observed were also very different for different types of discharge. For incomplete discharges, the highest frequency did not exceed 5 MHz, whereas in the final stage of the most powerful type III discharges, the signal was observed at frequencies higher than 150 MHz.

It is well known that electric discharges are accompanied by the generation of electromagnetic waves. The leader channel and the channel of the main discharge are filled with a highly conductive plasma. Such a plasma channel plays the role of an emitting antenna. Therefore, knowing the time behavior of the discharge current, we can find the spectrum of electromagnetic waves emitted during a discharge. In the dipole approximation, the intensity of the emitted electromagnetic radiation is proportional to the square of the Fourier transform of the time derivative of the discharge cur-

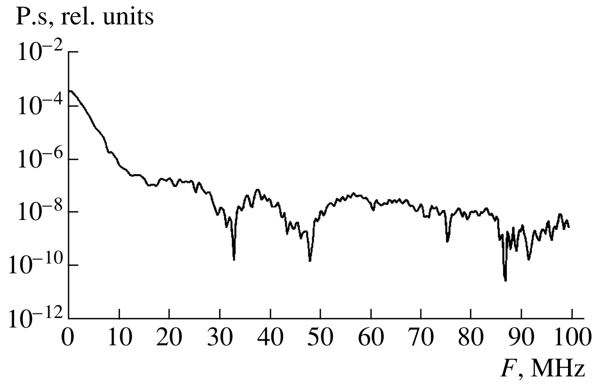


Fig. 6. Fourier spectrum of type III discharge.

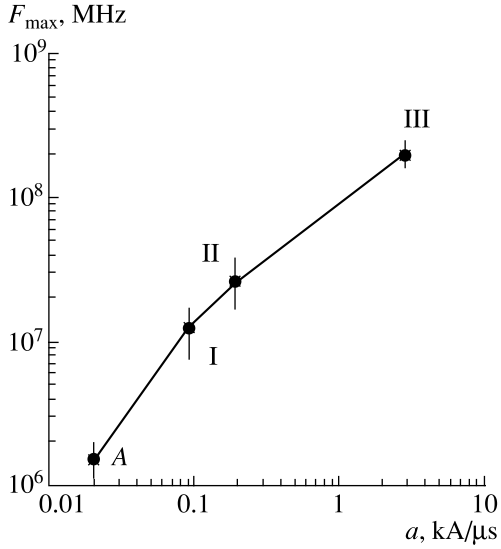


Fig. 7. Maximum characteristic frequency of the wavelet transform of the current signal for different types of discharge vs. current growth rate: (*A*) incomplete discharge and (I–III) different types of discharge.

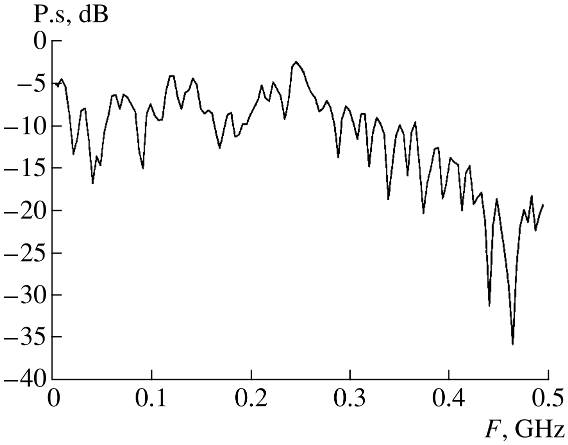


Fig. 8. Electromagnetic radiation spectrum generated by a type III discharge.

rent, i.e., to the square of the Fourier transform of the discharge current multiplied by the frequency squared. Figure 8 shows the Fourier spectrum of the electromagnetic radiation generated by a type III discharge. Thus, the above frequency characteristics of the discharge current provide information on the spectra of electromagnetic waves emitted during a discharge.

CONCLUSIONS

Our experimental studies of the development of a discharge between a charged aerosol cloud and a metal rod installed on a grounded plane have revealed three different types of discharge. It has been shown that the discharge parameters and the charge neutralized in the aerosol cloud depend substantially on the trajectory of the preceding leader discharge relative to the charged aerosol cloud, the place and angle at which the leader enters the cloud, and the probability of generating a counterpropagating discharge from the charged aerosol cloud.

Among other types of discharge between a charged aerosol cloud and a grounded electrode, type III discharges, which are characterized by the high brightness of the discharge channel, high current growth rate in the final stage (~1 kA/μs), and deep penetration into the charged aerosol cloud, are the closest to the main stage of natural lightning. A specific feature of powerful type III discharges, which best neutralize the cloud charge, is that such discharges enter the cloud at an angle of 40°–60° to its boundary, penetrate into the cloud to a depth of more than 20–30 cm, and neutralize up to 3 μC of the space charge of the aerosol cloud.

An analysis of wavelet and Fourier spectra of the discharge current has shown that the characteristic frequency ranges for different types of discharge between a charged water-aerosol cloud and a grounded electrode are close to one another, whereas the signal intensities at the characteristic frequencies are quite different. Wavelet analysis has shown that the signal intensity (i.e., the energy deposition rate in a discharge) is maximum at high frequencies. The more powerful the discharge (according to optical and current characteristics), the wider the characteristic frequency range and the higher the signal intensity in both the Fourier and wavelet spectra.

Among the discharges in a charged aerosol cloud, type III discharges are the most powerful. For such discharges, the maximum signal intensity (which is by one to two orders of magnitude higher than for other types of discharge) is achieved at frequencies of several hundred megahertz. Intense electromagnetic emission at such frequencies is typical of long streamer discharges, especially of negative streamers [14, 32]. A possible reason for the generation of an intense signal at such high frequencies in powerful discharges penetrating deep into the charged aerosol cloud is the development of a negative streamer corona propagating from the cloud boundary toward the upward leader approaching the cloud.

REFERENCES

1. E. M. Bazelyan and Yu. P. Raizer, *Lightning Physics and Lightning Protection* (Fizmatlit, Moscow, 2001).
2. E. R. Williams, C. M. Cooke, and K. A. Wright, J. Geophys. Res. **90**, 6059 (1985).
3. V. A. Rakov and M. A. Uman, *Lightning: Physics and Effects* (Cambridge Univ. Press, Cambridge, 2003).
4. E. J. Barreto, J. Geophys. Res. **74**, 6911 (1969).
5. T. Sugimoto, S. Tanaka, Y. Higashiyama, and K. Asano, IEEE Trans. Ind. Appl. **35**, 225 (1999).
6. A. Wada, A. Asakawa, and T. Shindo, in *Proceedings of the 26th International Conference on Lightning Protection, Crakow, 2002,* Vol. 1, pp. 3–7.

7. M. Miki, T. Shindo, V. A. Rakov, *et al.*, in *Proceedings of the 26th International Conference on Lightning Protection, Crakow, 2002*, Vol. 1, pp. 14–19.
8. V. A. Rakov, Bull. Am. Meteorol. Soc. **84**, 767 (2003).
9. R. J. Thomas, P. R. Krehbiel, W. Rison, *et al.*, Geophys. Res. Lett. **28**, 143 (2001).
10. V. A. Rakov and M. A. Uman, IEEE Trans. Electromagn. Compat. **38**, 156 (1996).
11. V. A. Rakov and M. A. Uman, IEEE Trans. Electromagn. Compat. **43**, 654 (2001).
12. X. M. Shao, P. R. Krehbiel, R. J. Thomas, *et al.*, J. Geophys. Res. **100**, 2749 (1995).
13. D. M. Le Vine, J. Geophys. Res. **85**, 4091 (1980).
14. D. A. Smith, X. M. Shao, D. N. Holden, *et al.*, J. Geophys. Res. **104**, 4189 (1999).
15. W. A. Lyons, T. E. Nelson, R. A. Armstrong, *et al.*, Bull. Am. Meteorol. Soc. **84**, 445 (2003).
16. C. J. Rodger, J. R. Wait, R. L. Dowden, *et al.*, J. Atmos. Sol.-Terr. Phys. **60**, 1193 (1998).
17. M. A. Uman and V. A. Rakov, Prog. Aerosp. Sci. **39**, 61 (2003).
18. V. J. Mazur, J. Geophys. Res. **94**, 3311 (1989).
19. A. G. Temnikov and A. V. Orlov, in *Proceedings of the 26th International Conference on Lightning Protection, Crakow, 2002*, Vol. 1, pp. 298–301.
20. A. G. Temnikov and A. V. Orlov, Elektrichestvo, No. 8, 18 (1996).
21. A. G. Temnikov, K. N. Firsov, L. M. Vasilyak, *et al.*, Teplofiz. Vys. Temp. **41**, 200 (2003).
22. V. N. Bolotov, S. E. Kolesnikov, and Yu. V. Tkach, Élektromagnitnye Yavleniya **2**, 335 (2001).
23. N. Takeuchi, K. I. Narita, and Y. Goto, J. Geophys. Res. **99**, 10751 (1994).
24. J.-I. Yano, M. W. Moncrieff, X. Wu, and M. Yamada, J. Atmos. Sci. **58**, 850 (2001).
25. J.-I. Yano, M. W. Moncrieff, and X. Wu, J. Atmos. Sci. **58**, 868 (2001).
26. M. Wada, T. Ushio, M. Kawada, *et al.*, Trans. Inst. Electr. Eng. Jpn., Part B **116**, 469 (1996).
27. V. V. Vinogradov and G. V. Litvinenko, in *Proceedings of the International Conference on Mathematical Methods in Electromagnetic Theory, Dunedin (New Zealand), 1998.*
28. A. Ageyev, A. Gandelli, and R. Zich, in *Proceedings of the 15th International Zurich Symposium and Technical Exhibition on Electromagnetic Compatibility, Zurich, 2003.*
29. A. Tungkanawanich, E. Y. Hamid, Z.-I. Kawasaki, and K. Matsuura, in *Proceedings of IEEE Power Engineering Society Meeting, OH, USA, 2001*, Vol. 1, pp. 263–268.
30. L. J. Lanzerotti, D. J. Thomson, C. G. Maclennon, *et al.*, J. Geophys. Res. **94**, 13221 (1989).
31. J. C. Wilett, J. C. Bailey, and E. P. Krider, J. Geophys. Res. **94**, 16255 (1989).
32. C. T. Rhodes, X. M. Shao, P. R. Krehbiel, *et al.*, J. Geophys. Res. **99**, 13059 (1994).

Translated by A. Sakharov

Technical Physics, Vol. 50, No. 7, 2005, pp. 876–880. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 60–64.
Original Russian Text Copyright © 2005 by Danchenko, Repin, Rep'ev.

GAS DISCHARGES, PLASMA

Experimental Study of the Discharge Glow Structure in the Initial Phase of a Nanosecond Breakdown of an Insulating Gap

E. G. Danchenko, P. B. Repin, and A. G. Rep'ev
All-Russia Research Institute of Experimental Physics, Russian Federal Nuclear Center,
Sarov, Nizhni Novgorod Oblast, 607188 Russia
e-mail: repin@ntc.vniief.ru

Received June 27, 2004; in final form, November 10, 2004

Abstract—The diffuse phase of high-voltage nanosecond atmospheric-pressure discharges in insulating gaps with a highly nonuniform electric field has been studied experimentally with the purpose of developing methods for protecting power supply equipment from natural and man-caused overvoltages. The structure of the discharge glow in 6- to 12-cm rod–plane gaps has been investigated for five configurations of the rod cathode. It is found that the discharge can exist in different forms and that the occurrence of one or another type of discharge is probabilistic in character. The electric characteristics of the discharge have been studied as functions of the electrode gap length for two types of cathode that provide the preferential development of volume and multichannel discharges. It is shown that the main factors governing the shape of the discharge glow are the distribution of the electric field near the cathode in the voltage growth phase and the dynamics of the accompanying discharge processes. *© 2005 Pleiades Publishing, Inc.*

Interest in studying high-voltage diffuse discharges in elevated-pressure gaseous media ($\leq 10^5$ Pa) in a highly nonuniform electric field stems, in particular, from the problem of protecting energy supply equipments. The amplitude and temporal characteristics of the overvoltages exerted on the elements of power supply systems and resulting in the breakdown of insulating gaps can vary in rather wide ranges. In particular, nanosecond overvoltages can arise under the action of both natural (e.g., in some stages of lightning [1]) and man-caused factors. The breakdown (short-circuiting) of a gap can be regarded as a stage completing the formation phase of a diffuse discharge. It is this phase that determines the possibility of transforming the discharge into an arc with a descending current–voltage characteristic. After such a transformation, the current continues to flow through the insulation gap after the end of the overvoltage pulse, being maintained by the operating voltage of the energy supply equipment.

Nanosecond discharges in ~1-cm-long gaps have been thoroughly investigated both experimentally and theoretically [2, 3]. However, in order to solve the problem of protecting energy supply equipments, it is necessary to study atmospheric-pressure discharges in air at gap lengths of ≥10 cm. Experimental studies in this field are rather limited in number because of the technical difficulties; this is why publications devoted to investigation of nanosecond breakdowns of long gaps are dominated by theoretical and computational works [4, 5]. At the same time, the strong dependence of the discharge characteristics on the external conditions (such as the gap geometry and the parameters of the applied voltage pulse) makes it necessary to carry out additional experiments. In this paper, we present results from experimental studies of nanosecond diffuse discharges developing in a rod–plane gap in atmospheric-pressure air. The appearance and the main electrical parameters of a diffuse discharge were investigated for different gap lengths h and different degrees of the electric field nonuniformity. The experiments were performed with five different configurations of the rod cathode. The gap length h was varied from 6 to 12 cm. The duration of the applied voltage pulse was limited so that the initial, slightly luminous phase of breakdown had not time to pass into a spark phase.

A schematic of the experimental setup is shown in Fig. 1. As a power supply (*1*), we used a BING-6 pulsed nanosecond 60-kV voltage generator [6] loaded on a Lewis transformer [7]. The transformer was assembled of six 17-m segments of a coaxial cable (with a wave impedance of 55 Ω), whose outputs were connected in

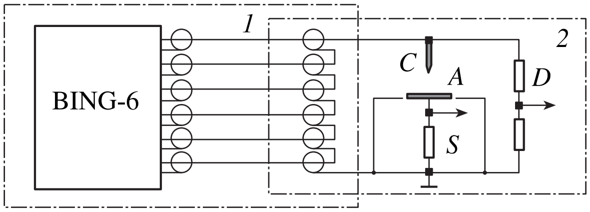


Fig. 1. Schematic of the experimental setup.

series in the upper part of the discharge chamber (*2*). The discharge chamber was a metal cylinder 60 cm in diameter and 60 cm in height. The maximum voltage was applied to the center of the chamber. 1-cm-diameter 15-cm-long rod cathode *C* was installed at the chamber axis so that the distance *h* between the cathode end and the anode could be varied. We used cathodes with axisymmetric 1-cm-diameter 4-cm-long stainless-steel tips of different shape (Fig. 2). The radius r_0 of tips *b*–*e* was ~0.2 mm (some experiments were performed with radii of r_0 = 0.01–1.6 mm). An 18-cm-diameter plane aluminum anode *A* was installed coaxially with the cathode at the bottom of the chamber. The anode was a part of the plane grounded basis of the chamber; this ensured the uniformity of the electric field near the plane electrode.

In our experiments, the discharge voltage was measured using resistive divider *D* and the discharge current was measured with the help of low-inductance shunt *S*. We also recorded the integral (over the gap length) radiation from the discharge in the wavelength range of 300–600 nm using a fast-operating photomultiplier. The electric signals were recorded by a digital oscilloscope with a passband of 500 MHz and time resolution of no worse than 0.5 ns for current and voltage signals and of ~2 ns for radiation pulses. The discharge glow was observed visually and was photographed (in all, about 1000 photographs were taken) by a reflex camera located at a distance of 0.4–1.0 m from the discharge axis (the cathode region of the discharge was photographed using a telescopic lens). As a photosensitive material, we used color and black-and-white films with a spectral sensitivity of up to 640 nm. The number and diameter of the discharge channels were determined by photographing and by the method of autographs [8].

It was found experimentally that, for the same geometry of the discharge gap and the same power supply parameters, the shape of the diffuse discharge varied from shot to shot. The shapes of the glow can be divided into three main types: multichannel, volume, and transitional. As a rule, a multichannel discharge (Fig. 3a) consists of four to sixty channels with nearly the same diameters and intensities. The channels are evenly distributed over the anode plane. In a volume discharge (Fig. 3d), the diameter and intensity of the central channel substantially exceed those of the remaining channels. By transitional discharges (Figs. 3b, 3c) we mean intermediate types of discharge that cannot be ascribed to any of the above types because of their very different (including asymmetric) and, as a rule, poorly reproducible shapes.

In terms of the above classification, the probability of the occurrence of one or another type of discharge is almost independent of the discharge gap length. At the same time, the geometric parameters of the glowing regions, as well as the emission characteristics of the discharge, depend substantially on *h*. The shorter the gap length *h*, the smaller (on average) the diameter of the discharge channels and the higher the intensity of the discharge glow. At large h, the volume discharge has the shape of two separate cones with a 7-cm-diameter common base in the central region of the discharge gap. As *h* decreases, the volume discharge acquires the shape of a cylinder with a diameter of ~1 cm. At *h* = 6 cm, among diffuse discharges, brightly glowing spark discharges with a small current-channel diameter are sometime observed.

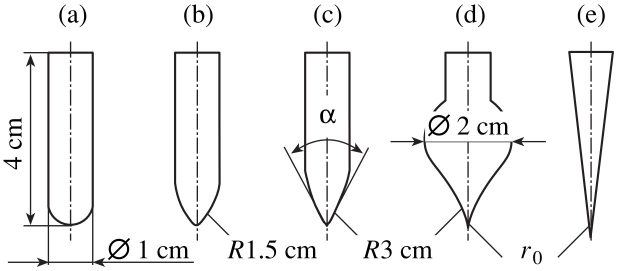


Fig. 2. Shapes of the cathode tips.

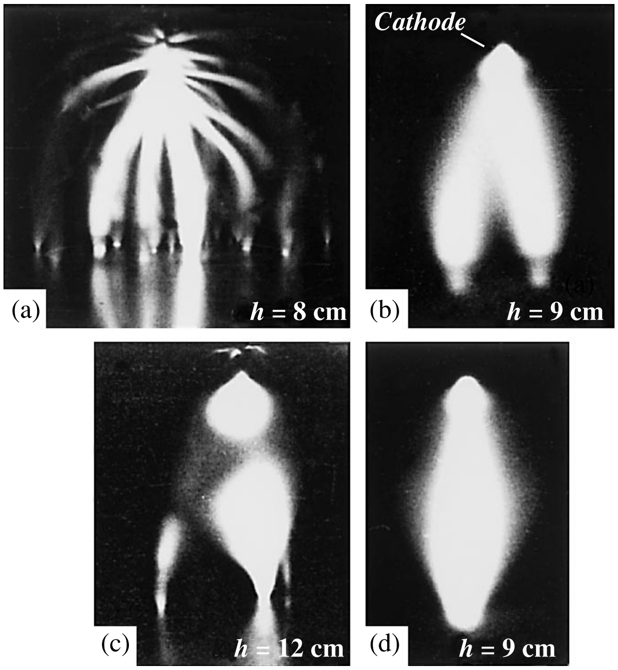


Fig. 3. Different types of discharge for a bullet-shaped cathode (Fig. 2c): (a) multichannel, (b, c) transitional, and (d) volume.

In spite of the absence of breakdown as it is understood in [5] (i.e., in spite of the absence of a highly conductive spark channel with a descending current–voltage characteristics), the process of discharge development at 6 < *h* ≤ 12 cm quite adequately reflects the prebreakdown stage of a discharge. This follows from a comparative analysis of the current and voltage waveforms of diffuse and spark discharges in short gaps (*h* = 5–6 cm). Under these conditions, the formation of a

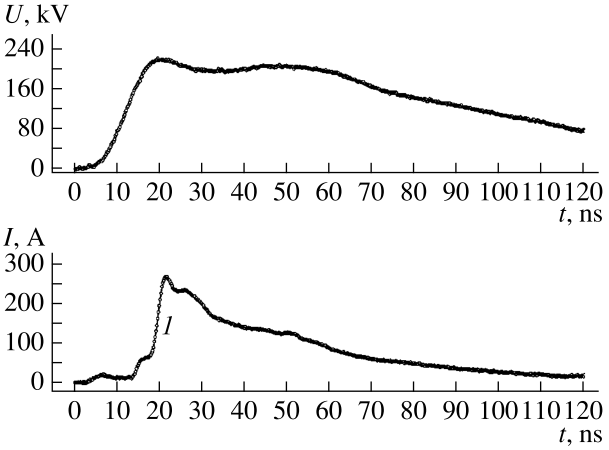


Fig. 4. Typical waveforms of the discharge voltage and current: (*1*) the initial phase of the conduction current.

highly conductive spark channel is usually observed either at the end of the voltage pulse or at its backswing. The segments of the current and voltage waveforms preceding the formation of a spark coincide (within the shot-to-shot scatter) with analogous segments for diffuse discharges. Moreover, variations in the frontal segments of the current and voltage waveforms correlate with the temporal parameters of a spark discharge.

Figure 4 shows typical waveforms of the discharge voltage and current. Variations in the shape of the discharge glow under the same formation conditions is accompanied by variations in the electric and emission parameters. Moreover, for discharges with the same shape of the glow, there is a statistical scatter in these parameters from shot to shot. The electric characteristics of the discharge were examined most thoroughly for hemispheric (Fig. 2a) and bullet-shaped (Fig. 2c) cathodes. It has been found that the increase in h from 6 to 12 cm leads to an increase in the peak voltage from 150 to 240 kV for a bullet-shaped cathode and from 140 to 230 kV for a hemispheric cathode. In this case, the discharge current amplitude exponentially drops from 800 to 150 A for a bullet-shaped cathode and to 300 A for a hemispheric cathode. At small gap lengths, the total duration of the current pulse is 170 ns. As h decreases, the total duration of the current pulse decreases to ~140 ns for a hemispheric cathode and to ~110 ns for a bullet-shaped cathode.

Probability of the occurrence of the different types of discharge as a function of α for r_0 = 0.2 mm and h = 10 cm

Cathode (Fig. 2)	α, deg	Type of discharge		
		multichannel	transitional	volume
a	180	0.7	0.2	0.1
b	90	0.2	0.3	0.5
c	60	0.1	0.2	0.7
d	30	0.1	0.3	0.6
e	15	0.1	0.2	0.7

At the instant of the maximum current, the discharge resistance at h = 6 cm is ~100 Ω for both types of cathode. It increases with increasing gap length and, at h = 12 cm, reaches 1.2 kΩ for a bullet-shaped cathode and 600 Ω for a hemispheric cathode. The time during which the gap is bridged by the discharge channels (this time is defined as a delay between the beginning of the voltage growth and the start of the conduction current, which corresponds to a sharp drop in the gap resistance) increases nearly linearly from 10 to 17 ns as h increases from 6 to 12 cm. Thus, the average propagation velocity of streamers in the gap depends slightly on h and is no less than 6×10^8 cm/s.

Special experiments carried out at a constant gap geometry did not reveal any external factors affecting the occurrence of different types of discharge glow. It was found that the probability of the occurrence of one or another type of discharge is independent of the electrode roughness; of small variations in the pressure, humidity, and temperature of air; and of the presence in the discharge gap of microparticles and long-lived products of the preceding discharges (singlet oxygen, ozone, etc.). A decrease in the total discharge current (by shunting the discharge gap) and the limitation of the current at the discharge periphery (by placing an annular dielectric plate on the anode) also did not affect the probability of the occurrence of one or another type of discharge. Hence, the structure of the discharge glow is determined exclusively by the internal mechanisms of the discharge.

It was also found that the probability of the occurrence of one or another type of discharge (the ratio of the number of discharges with a given shape of glow to the total number of discharges) depends on the shape of the cathode tip and is primarily determined by the cathode vertex angle α and the point radius r_0 (see table). It can be seen from the table that the probability of the occurrence of a multichannel discharge decreases with decreasing α, while the probability of the occurrence of a volume discharge increases. For a hemispheric cathode (Fig. 2a), multichannel discharges predominate, whereas for cathodes (Figs. 2b–2e) with r_0 = 0.2 mm, volume discharges prevail. An increase in the point radius of cathodes (Figs. 2b–2e) leads to an increase in the probability of the occurrence of transitional and multichannel discharges. As r_0 decreases to 10 μm, the probability of the occurrence of volume discharges increases by ~10–20%.

Apparently, the structure of the discharge glow forms in the initial phase of discharge due to the onset of instabilities of the cathode plasma sheath, as it occurs in high-voltage diffuse discharges between a plane and a wire parallel to this plane [9, 10]. This is confirmed by the similarity of the conditions under which a discharge forms (the voltage and current amplitudes, the voltage growth rate, the parameters of the working gas, and the discharge gap length) in our experiments and in [9, 10], by the similarity of the discharge glow structure (a multichannel discharge), and by the experimentally found dependence of the discharge glow structure on the cathode geometry.

In order to consider the influence of the cathode tip geometry on the shape of the discharge glow in more detail, we calculated the electric field strength around the cathode. The calculations were performed in real three-dimensional geometry with allowance for the influence of the zero potential of the discharge chamber wall on the electric field distribution in the electrode gap. We performed calculations for five cathode configurations shown in Fig. 2 at $h = 10$ cm and $r_0 = 0.2$ mm. For each configuration, the field was calculated at $\sim 10^6$ points. To improve the calculation accuracy, the density of the node points was increased near the critical regions of the cathode surface.

The results of calculations for hemispheric, bullet-shaped, and conical cathode tips (Figs. 2a, 2c, and 2e, respectively) are presented in Figs. 5 and 6. Calculations show that an increase in the angle α (a transition from a conical to a hemispheric cathode) leads to a substantial decrease in the maximum value of the electric field at a fixed applied voltage and to the smoothing of the field distribution along the cathode surface (Fig. 6). A the same time, Fig. 6 shows that, at small angles α, the length of the region in which the field strength is close to its maximum value for a hemispherical cathode increases. The experiments demonstrate that such a field is quite sufficient for a discharge to occur. By analogy with [9, 10], this would lead to an increase in the number of discharge channels and, thus, to an increase in the probability of the occurrence of a multichannel discharge with decreasing α, which contradicts the experimental data (see table).

In our experiments (in contrast to [9, 10]), the highly nonuniform distribution of the field along the cathode surface at small α, together with the finite value of the voltage growth rate, leads to a situation in which the electric field strength that is sufficient for the development of a discharge is reached at different times in different regions of the cathode surface, starting from the axial region. An analysis of the current waveforms and the relevant waveforms of discharge radiation in the UV and visible spectral regions shows that, in the gap bridging phase (the segment of the current waveform preceding to the conduction current; see Fig. 4), the charges begin to move (which apparently corresponds to the beginning of the formation of the discharge current structure) before the gap voltage reaches its maximum value. (More detailed data from studying the initial stage of discharge will be published in the near future.) The time delay between the beginning of the voltage growth and the instant at which the charges start moving depends on the cathode shape and the type of the forming discharge glow. For the types of discharge that are dominant for the cathodes in Figs. 2a, 2c, and 2e, the time delay is on average 14, 8, and 6 ns, respectively. The gap voltage at these instants is about 65, 20, and 15% of its maximum value. The calculated distributions of the electric field strength E_0 at these voltages are shown in Fig. 7.

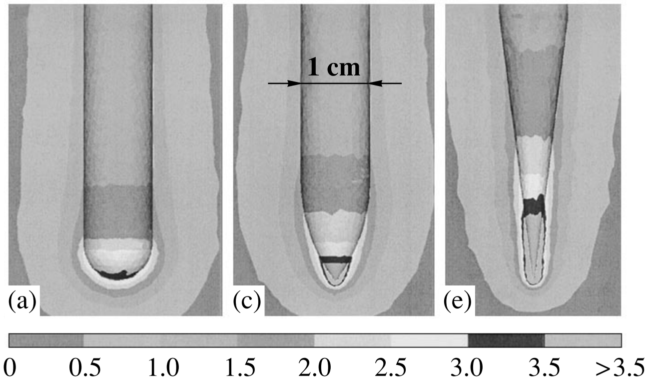


Fig. 5. Distribution of the electric field E near the ends of different cathodes (Figs. 2a, 2c, and 2e) at the same voltage of 200 kV.

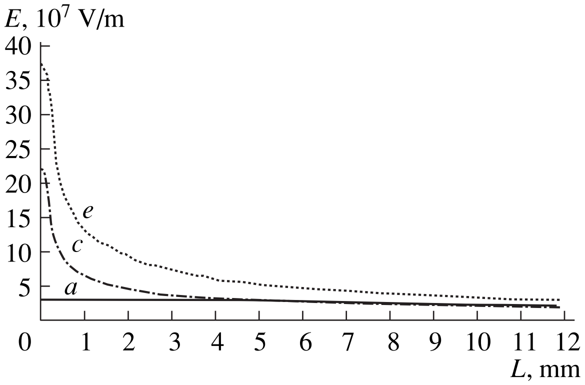


Fig. 6. Distribution of the electric field E along the surfaces of cathodes (Figs. 2a, 2c, and 2e) at a voltage of 200 kV.

It can be seen that the field distribution differs considerably from that shown in Fig. 5. At the instant at which the charges begin to move, the size of the region bounded by a surface of the same value of the electric field strength in the case of a hemispheric cathode is much larger than in case of cathodes with small angles α. A decrease in the radius r_0 of the cathode end leads to a further decrease in this size. If we assume that the formation of the discharge structure is governed by the instability of the cathode plasma sheath [9, 10], then, in accordance with the experimental results, we find that the probability of the occurrence of a multichannel discharge for a hemispheric cathode is higher than that for the other cathode configurations under study and, in contrast, the probability of the occurrence of a volume discharge is higher for cathodes with small α and r_0.

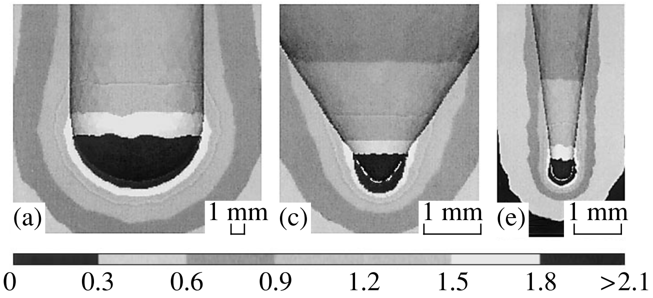


Fig. 7. Distribution of the electric field E_0 near the ends of the cathodes (Figs. 2a, 2c, and 2e) at the instant at which the charges begin to move.

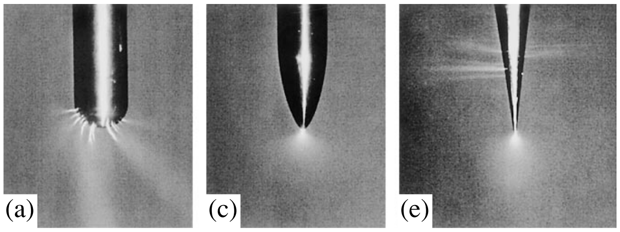


Fig. 8. Discharge glow near the cathodes (Figs. 2a, 2c, and 2e).

Moreover, the weak dependence of the field strength near the cathode on the electrode gap length leads to the weak influence of h on the probability of the occurrence of one or another type of discharge. Indeed, an increase in the gap voltage should appreciably change the discharge structure, the development of which starts when the charges begin to move. An increase in the voltage leads to an increase in the calculated size of the region where the field strength is sufficient for the beginning of breakdown; the actual distribution of the field, however, differs from that shown in Fig. 5 because of the screening and distortion of the field by the space charge in the gap. Moreover, perturbations that have already been formed on the surface of the cathode plasma sheath will develop with the highest growth rate. Their growth will suppress the development of new perturbations, at least at distances exceeding the characteristic distance between them. For cathodes with small angles α, only one discharge channel (a volume discharge) develops. This is probably explained by the too small size of the region in which the charges start moving before the development of perturbations of the cathode sheath. Moreover, the perturbation that is the forerunner of the discharge channel does not develop until the sheath reaches a certain size. This is confirmed by the fact that, as the angle α decreases, the calculated maximum electric field strength at the instant at which the charges begin to move increases from $\sim 2 \times 10^7$ V/m for a hemispheric cathode and to $\sim 5.5 \times 10^7$ V/m for a conical cathode.

Figure 8 shows the cathode region of the discharge. It can be seen that the channels of a multichannel discharge are evenly distributed over the surface of a hemispheric cathode. For a volume discharge with a bullet-shaped cathode, the peripheral channels are, as a rule, absent. With a conical cathode, for which the region of a strong field is more extended along the cathode surface (see Fig. 5), the channels are formed at a substantial distance from the main channel. In this case, however, the current flowing through the side channels is very low and they do not disturb the general pattern of the discharge glow, apparently, because of their late start and large length, which is determined by the field distribution in the discharge gap.

Thus, the observed features of a diffuse discharge can be explained using the model proposed in [9, 10], which describes the formation of the discharge structure in gaps with a highly nonuniform electric field in terms of the instability of the cathode plasma sheath. The main factors governing the shape of the discharge glow are the distribution of the electric field near the cathode surface in the voltage growth phase and the dynamics of the accompanying discharge processes.

REFERENCES

1. M. Uman, *Lightning Discharge* (Academic, London, 1987).
2. Yu. D. Korolev and G. A. Mesyats, *The Physics of Impulse Breakdown in Gases* (Nauka, Moscow, 1991; URO-Press, Yekaterinburg, 1998) [in Russian].
3. Yu. P. Raizer, *Gas Discharge Physics* (Nauka, Moscow, 1992; Springer-Verlag, Berlin, 1991).
4. A. E. Bazelyan and E. M. Bazelyan, Teplofiz. Vys. Temp. **31**, 867 (1993).
5. N. L. Aleksandrov, E. M. Bazelyan, N. A. Dyatko, and I. V. Kochetov, Fiz. Plazmy **24**, 587 (1998) [Plasma Phys. Rep. **24**, 541 (1998)].
6. B. G. Kudasov, S. S. Pavlov, V. A. Tananakin, *et al.*, in *Proceedings of the 11th IEEE International Pulsed Power Conference, Baltimore, 1997*, p. 1572.
7. I. A. D. Lewis and F. H. Wells, *Millimicrosecond Pulsed Techniques* (Pergamon, London, 1955; Inostrannaya Literatura, 1956).
8. P. B. Repin and A. G. Rep'ev, Zh. Tekh. Fiz. **71** (5), 128 (2001) [Tech. Phys. **46**, 632 (2001)].
9. A. I. Pavlovskii, M. A. Voinov, V. V. Gorokhov, *et al.*, Zh. Tekh. Fiz. **60** (1), 64 (1990) [Sov. Phys. Tech. Phys. **35**, 37 (1990)].
10. S. N. Buranov, V. V. Gorokhov, V. I. Karelin, *et al.*, Teplofiz. Vys. Temp. **29**, 383 (1991).

Translated by A. Sakharov

Technical Physics, Vol. 50, No. 7, 2005, pp. 881–885. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 65–69.

Formation of a Volume Discharge at a Subnanosecond Rise Time of the Voltage Pulse

I. D. Kostyrya*, V. M. Orlovskiĭ*, V. F. Tarasenko*,
A. N. Tkachev, and S. I. Yakovlenko****

** Institute of High-Current Electronics, Siberian Division, Russian Academy of Sciences,*
Akademicheskiĭ pr. 4, Tomsk, 634055 Russia

*** Prokhorov Institute of General Physics, Russian Academy of Sciences,*
ul. Vavilova 38, Moscow, 119991 Russia

e-mail: syakov@kapella.gri.ru

Received September 21, 2004

Abstract—A study is made of the formation of a volume discharge in atmospheric-pressure air in a nonuniform electric field without additional preionization. It is shown that the spatial distribution of the plasma glow between a plane and a spherical (as well as a point) electrode at a subnanosecond rise time of the high-voltage pulse is volumetric in character. The change of the voltage polarity does not qualitatively affect the character of the glow. The propagation of a spherical ionization wave in nitrogen is calculated in the drift-diffusion approximation. The fact that the character of the discharge glow is essentially independent of the voltage polarity is explained by the multiplication of the background electrons in the dense working gas. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

High-pressure pulsed volume discharges are widely used in different fields of science and technology. The results of investigations on the formation of such discharges are described in numerous reviews and monographs (see, e.g., [1–6]). There were also studies in which pulsed volume discharges in atmospheric-pressure air were formed without additional preionization [7–17]. In [7–16], the discharges were excited between a plane anode and a point cathode by applying nanosecond voltage pulses with an amplitude of ~100 kV and a rise time of ~1 ns. It was pointed out in [18–21] that, under certain conditions, the propagation of an ionization wave in a dense gas is determined not by the transport of electrons or photons, but rather by the multiplication of the background electrons, whose density rapidly increases in the external electric field preceding the breakdown [20, 21]. This mechanism for discharge propagation does not depend on the direction of the electric field. This allows one, in particular, to abandon the well-known photon hypothesis of streamer propagation [22]. The study of the propagation of an ionization wave in a dense gas is also important for understanding mechanisms for the generation of strong subnanosecond beams in gases at atmospheric pressure [8–10, 12, 15, 23, 24].

The aim of the present study is to investigate the influence of the voltage polarity on the formation of volume discharges excited in a dense gas by nanosecond voltage pulses with a subnanosecond rise time. It is shown that the spatial distribution of the plasma glow between a plane and a spherical electrode is essentially independent of the voltage polarity. Computation results demonstrating the effect of multiplication of the background electrons in a dense gas are also presented.

EXPERIMENTAL SETUP AND DIAGNOSTIC TECHNIQUES

Experiments were performed using a nanosecond pulsed generator from an ARINA X-ray unit [6]. The pulsed generator produced voltage pulses with an amplitude of up to 150 kV (the idle voltage at a high-resistance load). The FWHM of the voltage pulse was a few nanoseconds and depended on the load. The rise time of the voltage pulse was less than 1 ns. A specially designed insulator with an external diameter of 160 mm was installed at the generator output. With this insulator, no surface breakdown occurred even at a distance of 67 mm between the cathode and anode. A chamber made of 200-μm copper foil was attached to the generator. The inner diameter of the chamber was 160 mm. On the side opposite to the generator, the chamber terminated with a plane copper electrode. The cylindrical sidewall of the chamber had an 11 × 5 cm rectangular window. An aluminum sphere with a radius of 40 mm (cathode 1) or a 6-mm-diameted tube made of 50-μm steel foil (cathode 2) was used as a powered electrode. The experiments were conducted at an air pressure of 1 atm and electrode gap lengths of 22, 54, and 67 mm. In the course of experiments, we changed the voltage polarity, varied the distance between the electrodes, and used electrodes of various designs. The time-integrated

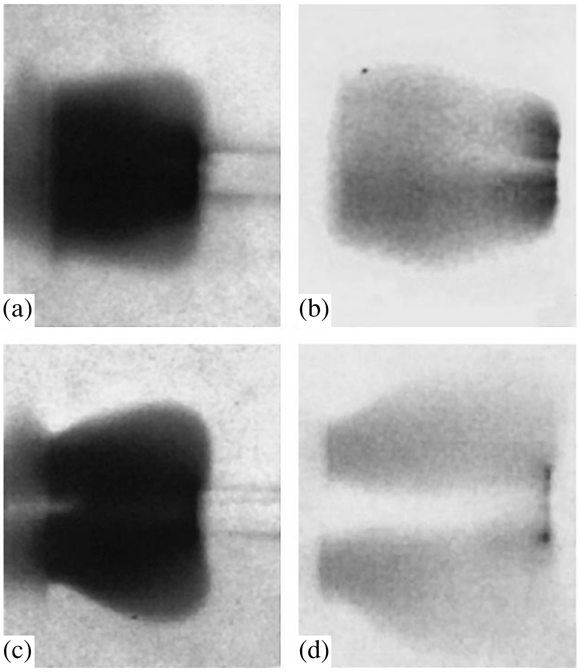


Fig. 1. Discharge glow observed in a 22-mm-long gap with cathode 2: (a, b) negative voltage polarity and (c, d) positive voltage polarity. Photographs (b) and (d) are obtained in one pulse, while photographs (a) and (c) are obtained by averaging over ten pulses. The powered electrode is on the right.

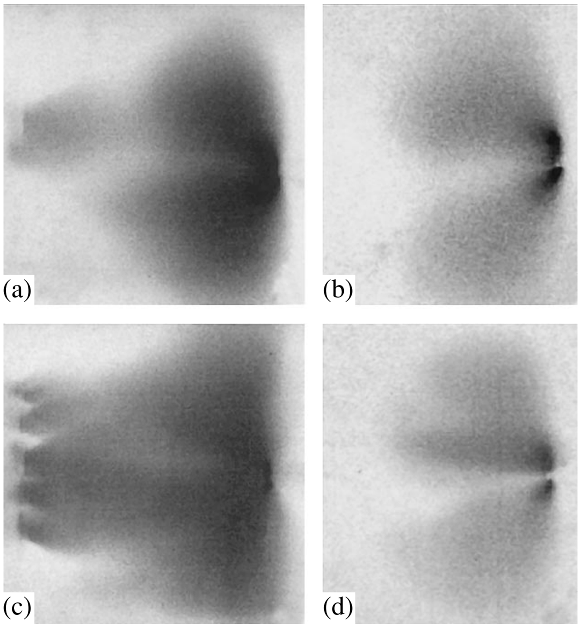


Fig. 2. The same as in Fig. 1, but for a 67-mm-long gap.

pattern of the discharge glow was recorded by a Zenith camera on an RF-3 film. We also measured the discharge current and discharge voltage pulses. A TDS-330 oscilloscope with a bandwidth of 300 MHz and resolution of 2.5 GS/s (2.5 dots per 1 ns) was used to record signals from the shunt and capacitive divider. X-ray emission was detected by the darkening of an RF-3 film, which was enclosed by black paper and was positioned at the side window of the chamber.

EXPERIMENTAL RESULTS

At the negative voltage polarity, a volume discharge, as in [7–16], was formed on the electrode with a small radius of curvature (Figs. 1a, 1b). Bright spots from which volume “jets” originated were observed only on the cathode. In [8], this type of discharge was called a volume discharge initiated by an electron avalanche beam (VDIEAB). An important feature of such a discharge with a point cathode [8–12] is the absence in the voltage signal of both the characteristic overvoltage peak and the phase of the fast voltage drop, which were observed in [1–3, 5, 6]. The discharge current starts to grow at the front of the voltage pulse and reaches its maximum essentially simultaneously with the voltage maximum, which corresponds to the voltage in the quasi-steady stage of the discharge.

With the reversed voltage polarity (i.e., with a plane cathode and a point anode), the volumetric character of the discharge was preserved (see Figs. 1c, 1d). In this case, however, the brightness of spots on the powered electrode decreased, bright spots on the plane cathode were absent, and the shape of the discharge glow was somewhat different. When the polarity of the powered electrode was positive, the size of the glowing area on the plane electrode was smaller and a minimum in the glow intensity was observed on the axis passing through the centers of the electrodes (see Fig. 1). Moreover, at the positive voltage polarity, the discharge glow expanded toward the insulator. This is clearly seen in a photograph of the plasma glow integrated over ten pulses (Fig. 1c). At a long discharge gap (>6 cm), regardless of the voltage polarity, the discharge had the shape of a diffuse torch on the powered electrode (Fig. 2); such a discharge can be classified as a pulsed corona discharge. In this case, the amplitude of the discharge current decreased significantly. When taking photographs averaged over ten pulses, discharges in which the plasma reached the cathode were sometimes observed (Figs. 2a, 2c); i.e., with the 67-mm discharge gap, it could be possible to obtain either a purely pulsed corona discharge or a discharge of an intermediate type.

The volumetric character of the discharge was also preserved when the electric field in the gap was significantly (by more than one order of magnitude) decreased. In this case, a metal sphere with a diameter of 40 mm was used as a powered electrode. Under these conditions, no spark channels were observed for either (positive or negative) voltage polarity. Bright spots were observed only on the powered electrode. Figure 3 shows the discharge glow in the case of a 22-mm discharge gap. Bright spots onto which the discharge closes through individual jets can be seen on the powered electrode. Two types of jets can be distinguished: type I jets bridge the discharge gap and have a higher brightness (a pulsed volume discharge), while type II jets start on the powered electrode and do not reach the opposite electrode (a pulsed corona discharge).

We believe that the main reason for the formation of a volume discharge under our experimental conditions (the subnanosecond rise time, large amplitude, and low duration of the voltage pulses) is preionization of the discharge gap by the runaway electrons and X radiation. Observations of X radiation and/or runaway electrons in the case of a point cathode were reported in many papers (see, e.g., [7–17] and references therein). For example, in optimizing a gas diode, the current of a subnanosecond electron beam formed in atmospheric-pressure air and withdrawn through a foil reached 240 A [16]. In our experiments, we also recorded pulsed X radiation when the polarity of the powered electrode was negative. No X radiation was observed when the polarity was positive.

ELECTRON MULTIPLICATION WAVE

Computation of a Spherical Multiplication Wave in Nitrogen

In [25], numerical simulations of electron multiplication in a cylindrical discharge gap filled with xenon were performed in order to determine conditions for the formation of an electron beam in a gas at atmospheric pressure. The simulations were performed in the drift-diffusion model proposed in [26, 27]. Here, we have developed a one-dimensional drift-diffusion model that allows one to qualitatively model ionization waves in nitrogen in a spherical (rather than cylindrical) sector.

Let us consider a plasma between spherical electrodes: $r_0 = 0.05$ cm $< r < r_1 = 3$ cm, where r_0 and r_1 are the radii of the external and internal electrodes, respectively. Plasma generation and electric field screening are described by the equations for momentum transfer and the continuity equations for ions and electrons, as well as Poisson's equation for the electric field. The dependences of the quantities entering in the drift-diffusion equations (the ionization frequency, drift velocities, and diffusion coefficients) on the electric field are specified using the data for nitrogen from [26]. The electrodes voltage is assumed to be a known function of time. For the computation results presented below, it was assumed that the voltage first grew linearly, reaching $U = 150$ kV at $t = 0.7$ ns, and then remained constant. The initial moderate background electrons density was set at $N_0 = 10^6$ cm^{-3}.

Our calculations show that the primary role in the plasma propagation is played by the multiplication of the background electrons (rather than by electron transport). Plasma starts to quickly approach the anode when the electrode voltage reaches a value of $U \sim$ 50 kV. The propagation velocity of the multiplication wave in this case turns out to be very high: at $U =$ 150 kV ($t \approx 1$ ns), the plasma approaches the anode with a velocity on the order of 3×10^9 cm/s (Fig. 4). This agrees with a simple formula for the propagation

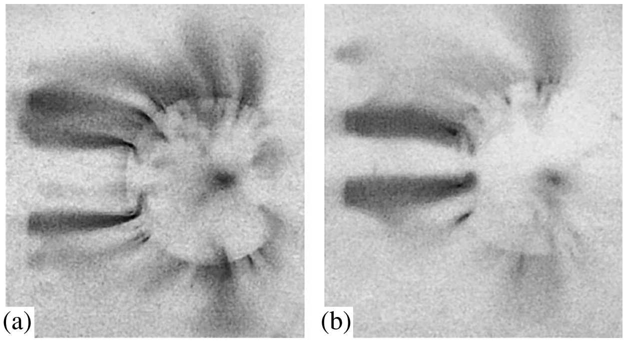


Fig. 3. Discharge glow observed in a 22-mm-long gap with cathode 1: (a) negative voltage polarity and (b) positive voltage polarity. The photographs are obtained in one pulse. The powered electrode is on the right.

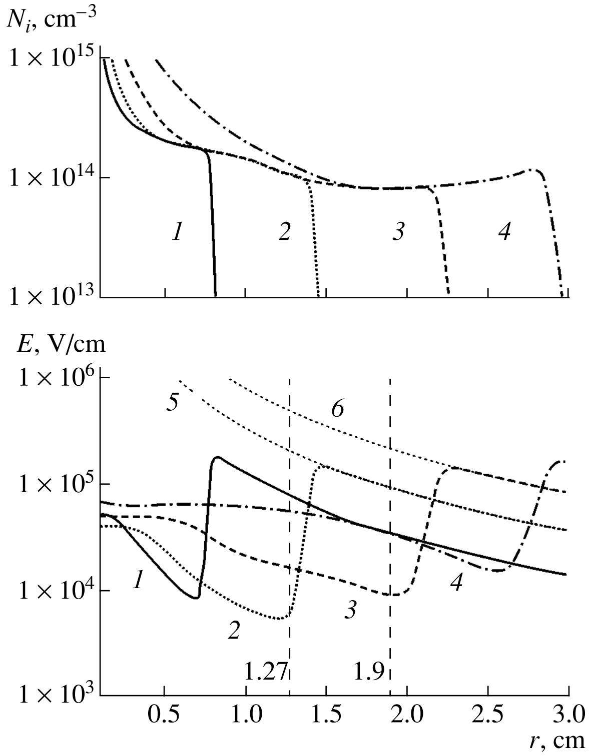


Fig. 4. Radial profiles of the (a) ion density N and (b) electric field E in the stage when the ionization wave approaches the anode: t = (*1*) 0.634, (*2*) 0.904, (*3*) 1.2, and (*4*) 1.38 ns. Curves *5* and *6* show the distribution of the field in empty space,

$$E(r, r_0, r_1) = U\frac{r_1 r_0}{(r_1 - r_0)}\frac{1}{r^2},$$

for $U = 150$ kV; $r_1 = 3$ cm; and $r_0 = 1.27$ and 1.9 mm, respectively.

velocity of an ionization wave obtained in [18–21]:

$$\nu = \frac{\nu_i r_0}{\zeta(E_0/p)},$$
$$\zeta(E_0/p) = 2\mathrm{Ln}\left(\frac{d\ln(u_{de}(E/p)\xi(E/p))}{d\ln(E/p)}\right)_{E/p\,=\,E_0/p}, \tag{1}$$

where ν_i is the ionization frequency, $E_0 = E(z(0))$ is the electric field at the front of the ionization wave, r_0 is the radius of the spherical surface approximating the front, p is the neutral gas pressure, $\mathrm{Ln} \equiv \ln(N_{\mathrm{cr}}/N_0)$, N_0 is the background plasma density, and N_{cr} is the critical plasma density at which the electric field is completely screened by the plasma. The ionization frequency can be written as the product $\nu_i = \alpha_i u_{de}$, where $\alpha_i(E, p) = p\xi(E/p)$ is the Townsend coefficient (with $\xi(E/p)$ being a function depending on the sort of gas) and $u_{de}(E/p)$ is the electrons drift velocity.

Let us assume, based on the data presented in Fig. 4, that $E_0 \approx 150$ kV/cm, $E_0/p \approx 200$ V(cm torr), $r_0 \approx 1.5$ cm, and $N_{\mathrm{cr}} = 10^{13}$ cm^{-3}. Taking into account the computation data from [28], we can write $\nu_i \approx 1.4 \times 10^{11}$ s^{-1} and $\zeta \approx 75$. Then, according to Eq. (1), we have $\nu \approx 3 \times 10^9$ cm/s.

Criterion for the Streamer Appearance

The condition $\alpha_i d > 20$ (where d is the distance between plane electrodes) is usually used as a criterion for the transformation of an avalanche into a streamer. When this condition is satisfied, the amount of charges produced in the avalanche is sufficient to provide the concentration of the electric field at its head.

It seems that this condition is insufficient, at least in the presence of background ionization. Actually, the volumetric multiplication of electrons can occur before the spark channel bridges the electrodes. A criterion for the transformation of a volume discharge into a spark can be written as

$$\frac{\nu(E_0)}{d} > \nu_i(E) \quad \text{or} \quad \frac{\nu_i(E_0)r_0}{\nu_i(E)\zeta(E_0)d} > 1,$$

where E is the electric field between the electrodes (this field is significantly lower than the field E_0 in the streamer head).

Therefore, if the electric field in the gap is sufficiently high to provide fast volumetric multiplication of electrons, the spark channel may not have time to appear. Of course, for this to occur, the field in the gap must increase sufficiently fast to avoid spark breakdown in the stage of the field growth. In other words, the rise time of the voltage applied to the electrodes must be sufficiently short.

In our opinion, there are sufficient reasons to assume that, in our experiments, we were dealing with a multiplication wave of background electrons. Since the character of the multiplication wave does not depend on the field direction, the shape of the discharge in Figs. 1–3 remains qualitatively the same when the voltage polarity changes its sign.

CONCLUSIONS

It has been shown that the breakdown of a relatively long (a few centimeters) discharge gap in atmospheric-pressure air by a high-voltage (~100 kV) pulse with a nanosecond duration and subnanosecond rise time in the absence of additional preionization leads to the formation of a volume discharge. In this case, the volumetric character of the discharge is preserved when the voltage polarity changes its sign, as well as when the electric field in the gap is decreased. It has also been shown that the volumetric character of the discharge at different voltage polarities can be explained by the formation of a multiplication wave of background electrons.

REFERENCES

1. G. A. Mesyats, Yu. I. Bychkov, and V. V. Kremnev, Usp. Fiz. Nauk **107**, 201 (1972) [Sov. Phys. Usp. **15**, 282 (1972)].
2. Yu. D. Korolev and G. A. Mesyats, *The Physics of Impulse Breakdown in Gases* (Nauka, Moscow, 1991; URO-Press, Yekaterinburg, 1998) [in Russian].
3. G. A. Mesyats, V. V. Osipov, and V. F. Tarasenko, *Pulsed Gas Lasers* (SPIE Optical Engineering Press, Bellingham, 1995).
4. Yu. P. Raizer, *Gas Discharge Physics* (Nauka, Moscow, 1992; Springer-Verlag, Berlin, 1991).
5. V. V. Osipov, Usp. Fiz. Nauk **170**, 225 (2000) [Phys. Usp. **43**, 221 (2000)].
6. G. A. Mesyats, *Pulsed Power* (Plenum, New York, 2004; Nauka, Moscow, 2004).
7. L. P. Babich, T. V. Loiko, and V. A. Tsukerman, Usp. Fiz. Nauk **160** (7), 49 (1990) [Sov. Phys. Usp. **33**, 521 (1990)].
8. V. F. Tarasenko, V. M. Orlovskii, and S. A. Shunailov, Izv. Vyssh. Uchebn. Zaved. Fiz. **46** (3), 94 (2003).
9. V. F. Tarasenko, S. I. Yakovlenko, V. M. Orlovskii, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **77**, 737 (2003) [JETP Lett. **77**, 611 (2003)].
10. V. F. Tarasenko, S. B. Alekseev, V. M. Orlovskii, *et al.*, Zh. Tekh. Fiz. **74** (8), 30 (2004) [Tech. Phys. **49**, 982 (2004)].
11. I. D. Kostyrya, V. S. Skakun, V. F. Tarasenko, and A. V. Fedenev, Zh. Tekh. Fiz. **74** (8), 35 (2004) [Tech. Phys. **49**, 987 (2004)].
12. V. F. Tarasenko, V. S. Skakun, I. D. Kostyrya, *et al.*, Laser Part. Beams **22**, 75 (2004).
13. V. F. Tarasenko, S. B. Alekseev, I. D. Kostyrya, *et al.*, Izv. Vyssh. Uchebn. Zaved. Fiz. **47** (2), 96 (2004).
14. I. D. Kostyrya, V. S. Skakun, V. F. Tarasenko, *et al.*, Pis’ma Zh. Tekh. Fiz. **30** (10), 31 (2004) [Tech. Phys. Lett. **30**, 411 (2004)].

15. S. B. Alekseev, V. P. Gubanov, V. M. Orlovskii, and V. F. Tarasenko, Pis'ma Zh. Tekh. Fiz. **30** (20), 35 (2004) [Tech. Phys. Lett. **30**, 859 (2004)].
16. I. D. Kostyrya and V. F. Tarasenko, Izv. Vyssh. Uchebn. Zaved. Fiz. **47** (12), 95 (2004).
17. P. B. Repin and A. G. Rep'ev, Zh. Tekh. Fiz. **74** (7), 33 (2004) [Tech. Phys. **49**, 839 (2004)].
18. S. I. Yakovlenko, Kratk. Soobshch. Fiz., No. 10, 27 (2003).
19. S. I. Yakovlenko, Pis'ma Zh. Tekh. Fiz. **30** (9), 12 (2004) [Tech. Phys. Lett. **30**, 354 (2004)].
20. S. I. Yakovlenko, Élektron. Zh. *Issledovano v Rossii*, No. 9, 86 (2004); http://zhurnal.ape.relarn.ru/articles/2004/009.pdf.
21. S. I. Yakovlenko, Zh. Tekh. Fiz. **74** (9), 47 (2004) [Tech. Phys. **49**, 1150 (2004)].
22. E. D. Lozanskii and O. B. Firsov, *Theory of Spark* (Atomizdat, Moscow, 1975) [in Russian].
23. V. F. Tarasenko and S. I. Yakovlenko, Usp. Fiz. Nauk **174**, 953 (2004) [Phys. Usp. **47**, 887 (2004)].
24. A. N. Tkachev and S. I. Yakovlenko, Cent. Eur. J. Phys. **2** (4); www.cesj.com/physics.html.
25. V. F. Tarasenko, S. I. Yakovlenko, V. M. Orlovskii, and A. N. Tkachev, Pis'ma Zh. Tekh. Fiz. **30** (8), 68 (2004) [Tech. Phys. Lett. **30**, 335 (2004)].
26. A. N. Tkachev and S. I. Yakovlenko, Laser Phys. **12** (7), 1022 (2002).
27. A. N. Tkachev and S. I. Yakovlenko, Laser Phys. **13** (11), 1345 (2003).
28. A. N. Tkachev and S. I. Yakovlenko, Pis'ma Zh. Tekh. Fiz. **30** (7), 14 (2004) [Tech. Phys. Lett. **30**, 265 (2004)].

Translated by B. Chernyavskiĭ

Technical Physics, Vol. 50, No. 7, 2005, pp. 886–890. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 70–74.
Original Russian Text Copyright © 2005 by Abramov, Abramova, Surovtsev.

SOLID-STATE ELECTRONICS

Dimensional Etching of Silicon and Silicon Dioxide in a Localized Gas Discharge

A. V. Abramov*, E. A. Abramova, and I. S. Surovtsev****

** Voronezh State University, Voronezh, 394693 Russia*

e-mail: Abramov@niif.vsu.ru

*** Voronezh State Architectural and Building University, Voronezh, 394006 Russia*

Received August 13, 2004

Abstract—The feasibility of dimensional maskless etching in a localized gas discharge is studied. The shape of the discharge is specified by the patterned surface of one of the electrodes. Patterning of SiO_2-on-Si films using this technique with a resolution of ≈ 5 μm is reported for the first time. The basic resolution-controlling parameters of this process are determined. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

In [1], we showed that a localized gas discharge (LGD) can be used for maskless patterning of the surfaces of various materials. Here, an rf capacitive discharge itself serves as an “engraver.” In this case, the negative image of a desired pattern is preapplied on a planar electrode located over a surface to be processed. We found that, at a certain combination of gas pressure P and discharge gap length L, the discharge configuration may completely follow the arrangement of features on the electrode surface that stand out toward the target; namely, in a chemically active discharge environment, etching occurs only under these features. Since the gap is extremely narrow (L may be less than 100 μm), the discharge power density may be raised by four to five times compared with that in conventional plasma-chemical etchers without additional power consumption. Therefore, the etching rate is also higher by several orders of magnitude. Such a discharge can be used to advantage for both masked and maskless rapid etching of materials surfaces. In this work, we consider the feasibility of applying an LGD in dimensional maskless etching of Si and SiO_2 with the aim to find the main factors governing the resolution of this process.

EXPERIMENTAL

The discharge chamber was evacuated with an NVR-5D pump. A 13.56-MHz discharge-initiating voltage was applied from a UV-1 generator, which makes it possible to control microwave power P_{rf} absorbed by a load. The topological features of the electrode surface and the profiles of etch grooves were measured with an MII-4 interferometer. During the experiments, the gas pressure was varied in the range 10^4–10^5 Pa. A homogeneous glow discharge under the prominent parts of the electrode surface was sustained at $P \times L$ = (2–10) Pa m. This range of $P \times L$ is near a minimum in the Paschen curve and corresponds to the conditions of existence of a normal glow discharge. At effective voltage drop U_{rf} = 150–300 V across the discharge, the LGD power density may reach 10 kW/cm^3. Gas overheating and so glow-to-arc changeover were eliminated because of efficient heat removal (owing to small L). The values of $P \times L$ mentioned above provide the relationship $L \approx 10\lambda_e$, where λ_e is the electron mean free path. Note for comparison that the discharge power density in standard plasma-chemical etchers does not exceed 1 W/cm^3 [2].

RESULTS AND DISCUSSION

Etching mechanism. Our experiments show that LGD etching of Si and SiO_2 is basically of a chemical nature, like conventional plasma-chemical etching, as indicated by, e.g., the selectivity of the process: the Si-to-SiO_2 etch rate ratio changes from 6 to 0.5 when the etching gas changes from SF_6 to CF_4. Another fact in favor of the chemical nature of the etching process is that etching of Si through an aluminum mask in the same gases is isotropic. Experiments on LGD etching of Si and SiO_2 in argon demonstrate that the physical sputtering of these materials can be neglected because of a low ion energy, which is less than several tens of electron volts. However, both experimental and simulation data indicate that ion bombardment contributes to surface activation and cleaning [3] and, thus, plays an important role in the LGD etching of Si and especially SiO_2.

It should be noted that the etching process under study has considerable distinctions from plasma-chemical etching and reactive ion-plasma etching. First, this process offers an extremely high etching rate due to a high concentration of reactive particles in the discharge. There are also other important, while not so apparent, differences stemming from the specific conditions of LGD existence. For example, the conventional methods of etching are based on the α-shape of the rf discharge, while the LGD is initiated in the transition regime from the rf γ discharge to the ac discharge. Its shape depends mainly on the relationships between the voltage frequency, gas pressure, and electrode distance [4]. For an electric field strength of 10^4–10^5 V/cm, which is characteristic of the LGD, the electron drift path length is much greater than L. Under such conditions, the discharge becomes impulsive and a cathodic layer forms at either of the electrodes alternatively in either half-cycle of the voltage oscillations. The surface being processed is subjected to ion bombardment at the times the glow is at the prominent parts of the electrode placed oppositely. The ions are accelerated normally to the surface being processed, which improves the accuracy of pattern transfer from the electrode surface. These features of the discharge favor etching localization. Note that the Debye screening length is comparable to L under these conditions; therefore, a stripped plasma cannot form.

Etching uniformity. A necessary condition for etching to be uniform is strict parallelism between a surface to be processed and the plane of the patterned electrode. Figure 1 shows the profiles of an etch pit and a cylindrical projection on the electrode that is coaxial with the pit. The profile of the pit is seen to copy that of the projection. The smaller the initial electrode–surface distance, the larger etch depth H. In our case, the initial scatter in these distances over the discharge's cross section was about 2 μm. The effect of scatter in L across the discharge on the etch uniformity is likely to be due to two factors. First, the discharge may not entirely cover the electrode surface at once, initially forming at places with smaller L; correspondingly, the etch time is longer at these places. Second, as the discharge current increases and the discharge occupies the whole electrode surface, the electric field strength, charged particles density, and, hence, reactive particle concentration are higher where L is shorter. Thus, to achieve the etch uniformity required, it is necessary to ensure the constancy of L across the discharge area and provide conditions for the gas composition in the discharge to be uniform. The fact that, under certain conditions, the discharge is initiated at sites where the gap between the patterned electrode and surface being processed is the shortest can be used for smoothing out the surface. This property of the etching technique considered is another specific feature distinguishing it from conventional methods.

Resolution. The most important characteristic of any process of dimensional etching is its resolution. For the conventional methods of masked plasma-chemical etching, this parameter is restricted, first, by a finite value of anisotropy coefficient A, which is equal to the ratio of H to the undercut. Second, etching may damage the mask via chemical interaction with plasma particles and physical sputtering. In our case, A can be defined as the ratio of H to Δl, where Δl is the distance from the boundary of the projection of an electrode feature onto the surface being processed to the boundary of the etch area [1]. One should bear in mind that Δl takes a final value virtually at the instant of discharge initiation and depends on time and H only slightly (at $H \ll L$). Therefore, A increases with etch depth and should be considered only hand in hand with H.

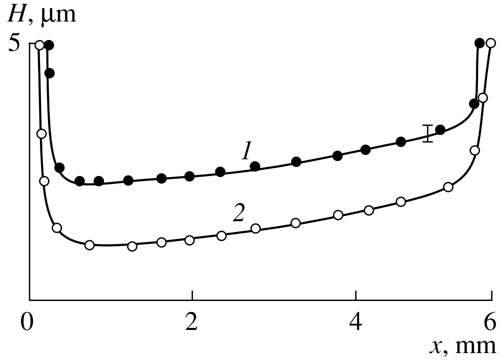


Fig. 1. Profiles of (*1*) the prominent feature on the electrode and (*2*) the pit etched in Si under the feature. The SF_6 atmosphere, $P = 2 \times 10^4$ Pa, $P_{rf} = 30$ W, $U_{rf} = 200$ V, $L = 300$ μm, and $t = 10$ s.

Our experiments show that the discharge can be localized at prominent features even if the depth of the electrode pattern is well below L. In one experiment, as an rf electrode, we used a silicon wafer covered by a thermally evaporated 1-μm-thick Al film with regularly spaced holes 60 (Fig. 2a) and 40 μm (Fig. 2b) in diameter obtained by etching. In the former case, the hole spacing was ≈10 μm; in the latter, ≈3 μm. On the other (grounded) electrode, we placed an oxidized silicon wafer (the SiO_2 thickness was 0.3 μm) so that the distance between the surfaces of these two wafers facing each other (L) was 85 ± 0.5 μm. A discharge initiated between the wafers at $P = 3 \times 10^4$ Pa in the SF_6 atmosphere for 1 s made it possible to etch patterns (Figs. 2c, 2d) in the SiO_2 film that were negative relative to the pattern in the Al film. This is not surprising, since the discharge was sustained only under the aluminum-covered surface areas on the wafer. Under each of the holes 60 μm in diameter, there were separate intact SiO_2 areas slightly less than 50 μm in diameter (Fig. 2c). It is easy to see that the undercut in this case is ≈5 μm; i.e., $\Delta l \ll L$. This is because the boundaries of SiO_2 etching are specified largely by the ion flux density distribution over the surface being processed and also because the ion flux is incident normally to this surface. In general, LGD charged components are distributed over a smaller volume than uncharged reactive particles, although the electron mean free path is longer. The point is that the electric field strength decreases with increasing distance from a protrusion on the electrode; hence, so do the electron energy and the molecule ion-

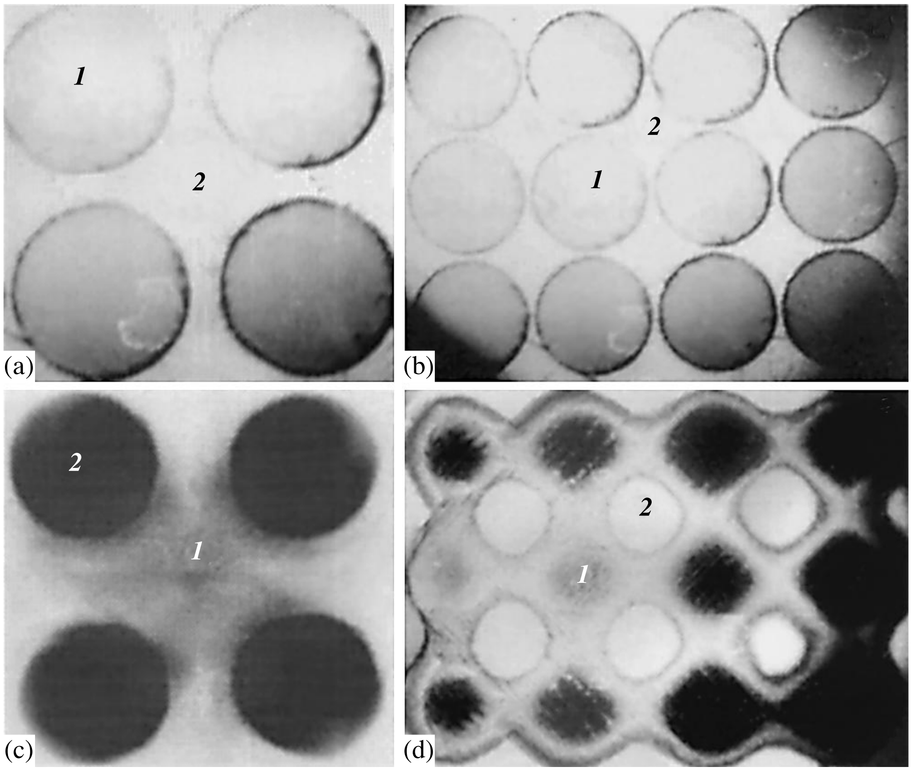


Fig. 2. (a, b) Patterned areas on the electrode ((*1*) Si and (*2*) Al) and (c, d) the patterns etched off in the SiO_2-on-Si film under these areas ((*1*) Si and (*2*) SiO_2).

ization frequency. Moreover, as the electron energy decreases, the probability of electron attachment to molecules grows. The attachment is, in fact, the basic obstacle to electron runaway from the place of their generation. Comparing Figs. 2b and 2d shows that, when the hole spacing in the Al film is 3 μm, the etching resolution is insufficient to produce an adequate negative image of the electrode pattern. Under the narrow regions of the aluminum film, the SiO_2 layer was not etched off completely. Here, the flux densities of reactive particles and ions incident on the surface being processed are likely to be insufficient because of scattering.

It should be noted that uncharged reactive particles may leave the discharge region and cause radical etching. However, nonvolatile reaction products arising in

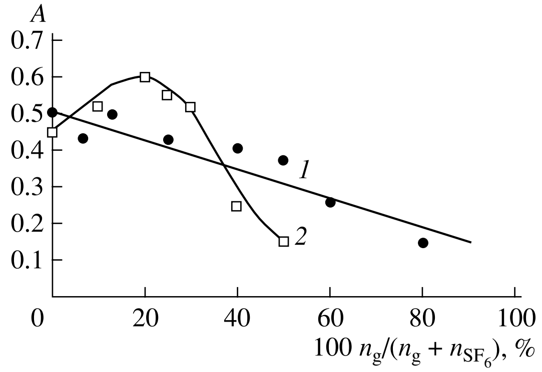


Fig. 3. Anisotropy of oxidized silicon etching vs. (*1*) Ar and (*2*) O_2 content in SF_6. $P_{rf} = 20$ W, $U_{rf} = 175$ V, $L = 100$ μm, $H = 20$ μm, and $P = 3 \times 10^4$ Pa.

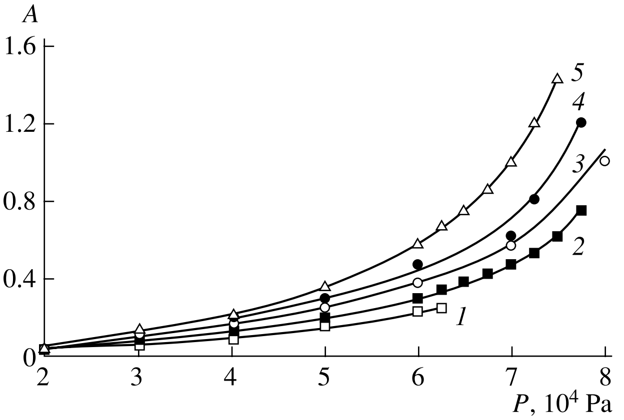


Fig. 4. Anisotropy of oxidized Si etching vs. the pressure of the gas mixture. $U_{rf} = 170$ V, $P_{rf} = 25$ W, $L = 100$ μm, and $H = 10$ μm. (*1*) CF_4, (*2*) CF_4 + 10% Ar, (*3*) CF_4 + 20% Ar, (*4*) CF_4 + 30% Ar, and (*5*) CF_4 + 40% Ar.

the discharge region are deposited on the target material and suppress this process. The masking effect of these products increases with distance from a prominent feature on the electrode, since the energy and density of the charged particle flux decline and become insufficient to activate and clean the surface.

The profiles of grooves etched off under the same conditions in the oxidized and unoxidized silicon shows that, in the latter case, *A* is several times higher. Remarkably, for the oxidized Si, Δl is lower and *H* is higher than for the uncovered Si surface. In these experiments, the etch time of Si was much longer than the time of SiO_2 removal. The fact that the etch grooves in the oxidized Si are narrower than in the oxidized silicon is explained by the masking effect of SiO_2. The oxide cannot be etched off if incident ions are not sufficiently intense. However, the SiO_2 etching area is smaller; accordingly, the concentration of reactive particles over the SiO_2 and, hence, the etching rate are higher. Furthermore, SiO_2 is an insulator and the discharge used in this work is an ac discharge. Therefore, the discharge current, being a conduction current, is likely to be confined toward the SiO_2-free region. This circumstance is bound to facilitate discharge localization and raise the etching rate.

We studied the dependence of *A* on the percentage of argon or oxygen additions to SF_6 and CF_4 at etching of 10-mm-long grooves on oxidized Si (the SiO_2 thickness was 0.72 μm) using a nickel electrode 100 μm wide. The etching anisotropy first increases with Ar content in CF_4 (up to 50%) and then decreases. A similar dependence is observed for an SF_6–argon mixture; however, here *A* reaches a maximum at about 20% Ar (Fig. 3). This result can be explained by the fact that argon, being added to a fluorine-containing gas, raises the fraction of positive ions in the net flux of positive ions and reactive particles. Therefore, the role of ion bombardment as a contributor to etching anisotropy increases. When the argon content in the mixture is higher than 20%, *A* begins to decrease, since the discharge expands as a result of the lowering probability of electron attachment. With oxygen added to SF_6 or CF_4, the etching anisotropy decreases virtually linearly because of long-lived oxygen-containing compounds, such as COF_2, OF_2, etc., forming in the discharge. The dissociation energies of these compounds are well below those of initial molecules of fluorine-containing gases [5, 6]. These compounds may dissociate much farther from a protrusion on the electrode, i.e., in those areas where the electric field strength and the electron temperature are lower. Moreover, these compounds may dissociate on the silicon surface as a result of dissociative chemisorption. Oxygen added to fluorine-containing gases may also extend the lifetime of reactive particles in these gases, oxidizing the particles and, thus, decelerating their reaction with other particles [7].

Figure 4 shows the dependence of *A* on *P* when a SiO_2 film 0.3 μm thick was etched at $H = 10$ μm. It is seen that the anisotropy increases with *P*. Earlier, the same dependence was observed for etching of unoxidized Si [1]. Such behavior can be accounted for by the fact that the electron concentration in the discharge depends on the pressure only slightly and the electron diffusion coefficient decreases in inverse proportion to *P*. In addition, as *P* grows, the electron energy decreases, which raises the probability of their attachment to molecules and also favors discharge localization in a smaller volume. It should be noted that the etching anisotropy can be increased by increasing *P* until the glow discharge transforms into an arc discharge. Since an increase in the pressure at a fixed *L* results in an increase in the breakdown voltage, it is necessary to decrease the electrode spacing as *P* increases.

Under the same conditions of Si etching by a localized discharge in SF_6 or CF_4, Δl in SF_6 is approximately two to three times smaller owing to electron attachment.

CONCLUSIONS

Our investigation shows that an LGD can be used for maskless dimensional etching of materials, which, basically, makes it possible to exclude the lithography stage from the fabrication process. The etching resolution is found to rise with increasing gas pressure and the electronegativity of the gas, as well as with shrinking the discharge gap. The resolution can also be substantially improved by optimizing the role of ion bombardment in etching. This can be done by appropriately choosing the composition of the plasma-forming gas with regard to the physicochemical properties of a material to be processed. Thus, the basic factors influencing the resolution of LGD etching are revealed. Taking into account that the characteristics of this process can be improved further and also its extremely high productivity, we believe that LGD etching may find application in various industries.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education of the Russian Federation in the framework of the program “Research of Higher School in Selected Areas of Science and Engineering,” project no. 04.01.18, subprogram “Electronics.”

REFERENCES

1. A. V. Abramov, E. A. Abramova, and I. S. Surovtsev, Pis’ma Zh. Tekh. Fiz. **27** (2), 45 (2001) [Tech. Phys. Lett. **27**, 108 (2001)].

2. B. S. Danilin and V. Yu. Kireev, *Low-Temperature Plasma for Etching and Cleaning of Metals* (Énergoatomizdat, Moscow, 1987), p. 264 [in Russian].
3. A. V. Abramov, E. A. Abramova, Yu. I. Dikarev, *et al.*, *Mathematical Modeling of Engineering Systems: A Collection of Scientific Works* (VGTA, Voronezh, 2002), No. 5, pp. 96–100.
4. Yu. P. Raizer, M. N. Shneider, and N. A. Yatsenko, *Radio-Frequency Capacitive Discharges* (Izd. MFTI, Moscow, 1995; CRC, Boca Raton, 1995).
5. C. L. Mogab, A. C. Adams, and D. Flamm, J. Appl. Phys. **49**, 3796 (1978).
6. Yu. I. Dikarev and V. I. Svettsov, Elektron. Tekh., Ser. Mater., No. 10, 41 (1983).
7. L. N. Vladimirova, Yu. I. Dikarev, and S. M. Tsvetkov, Kondensirovannye Sredy i Mezhfaznye Granitsy **4**, 79 (2002).

Translated by K. Shakhlevich

Technical Physics, Vol. 50, No. 7, 2005, pp. 891–895. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 75–79.
Original Russian Text Copyright © 2005 by Boikov, Danilov.

SOLID-STATE ELECTRONICS

Capacitance of the $La_{0.67}Sr_{0.33}MnO_3/(Ba,Sr)TiO_3$ Interface Induced by Electric Field Penetration into the Manganite Electrode

Yu. A. Boikov and V. A. Danilov
Ioffe Physicotechnical Institute, Russian Academy of Sciences,
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
e-mail: yu.boikov@mail.ioffe.ru

Received October 13, 2004

Abstract—(001)$La_{0.67}Sr_{0.33}MnO_3$/(001)$(Ba_xSr_{1-x}TiO_3$/(001)$La_{0.67}Sr_{0.33}MnO_3$ (x = 0–0.25) three-layer heterostructures are grown by laser evaporation on (001)$La_{0.3}Sr_{0.7}Al_{0.65}Ta_{0.35}O_3$ single-crystalline substrates. In a wide temperature range (≈150 K), effective permittivity ε of (1000 nm)$Ba_{0.25}Sr_{0.75}TiO_3$ and (1000 nm)$SrTiO_3$ films grown obeys the relationship $\varepsilon \sim (T - T_{CW})^{-1}$, where T_{CW} is the Curie–Weiss temperature for related bulk crystals. Using experimental dependences $\varepsilon(T)$, the capacitance of the (001)$La_{0.67}Sr_{0.33}MnO_3$/(001)$Ba_xSr_{1-x}TiO_3$ and (001)$La_{0.67}Sr_{0.33}MnO_3$/(001)$SrTiO_3$ interfaces, which is due to electric field penetration into the manganite electrode, is estimated ($C_{int} \approx 4\ \mu F/cm^2$). At bias voltages of ±2.5 V, the change in the permittivity of the STO and BSTO films in the heterostructures studied reaches 25 and 45%, respectively. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Growth and properties of multilayer heterostructures consisting of thin films of conducting perovskite-like manganites $La_{1-y}(Sr,Ca)_yMnO_3$ spaced by epitaxial layers of nonlinear ferroelectrics $Ba_xSr_{1-x}TiO_3$ have recently become a subject of extensive research [1, 2]. These heterostructures are promising for capacitors controlled by an electric [3] or magnetic [4] field. At temperatures below Curie temperature T_C, the concentration of localized holes in $La_{0.67}(Sr,Ca)_{0.33}MnO_3$ solid solutions reaches 6×10^{21} cm^{-3} and the electrical conductivity is comparable to that of “impure” metals. The Curie temperature of $La_{0.67}Sr_{0.33}MnO_3$ (LSMO) is roughly equal to 360 K [5], and the room-temperature conductivity of this compound is rather high. The inherent similarity of the $La_{0.67}(Sr,Ca)_{0.33}MnO_3$ and $Ba_xSr_{1-x}TiO_3$ structures, along with their good chemical compatibility (at $T < 800$°C), makes it possible to grow perfect epitaxial heterostructures where manganite layers are in contact with insulating (ferroelectric) spacings.

At present, available data for the electronic parameters of interfaces between conducting oxide electrodes and ferroelectric spacers in multilayer epitaxial systems are sparse and fragmentary [6, 7]. Also, very little is known about the effect of mechanical stresses or nonstoichiometry on the properties of these interfaces.

In this work, we study the structure and permittivity of $SrTiO_3$ (STO) and $Ba_{0.25}Sr_{0.75}TiO_3$ (BSTO) films in parallel-plate capacitor heterostructures with LSMO electrodes and estimate the capacitance induced by electric field penetration into the manganite electrode.

EXPERIMENTAL

Manganite and ferroelectric layers in the order LSMO $\longrightarrow$ BSTO(STO) $\longrightarrow$ LSMO were grown on a (001)$La_{0.3}Sr_{0.7}Al_{0.65}Ta_{0.35}O_3$ (LSATO) substrate by laser evaporation (ComPex 205 KrF 248-nm laser, pulse width τ = 30 ns). The intermediate BSTO (STO) layer was d = 1000 nm thick, and the thickness of the upper and lower manganite layers was d_1 = 120 nm. During growth of three-layer manganite–ferroelectric–manganite heterostructures, the substrate temperature was kept at 770°C and the oxygen pressure in the growth chamber was about 0.3 mbar.

The phase composition of the LSMO/STO/LSMO and LSMO/BSTO/LSMO structures, lattice parameters in the manganite and ferroelectric layers, and their orientation were determined by X-ray diffraction (Philips X’pert MRD diffractometer, $\omega/2\theta$ and ϕ scannings). The lattice parameters of the films were measured both in and normally to the substrate plane by visualizing the $\omega/2\theta$ scannings when the incident and reflected X-ray beams lay in the plane normal to the (001) or (101) plane of the LSATO.

The surface morphology was examined with a NanoScope-IIIa atomic force microscope.

Square contact pads ($S = 200 \times 200\ \mu$m) on the upper manganite layer and contact vias in the insulating film (to provide the contact with the common (lower) LSMO electrode) were made by photolithography and Ar ion etching (500 V, 0.2 mA).

1063-7842/05/5007-0891$26.00 © 2005 Pleiades Publishing, Inc.

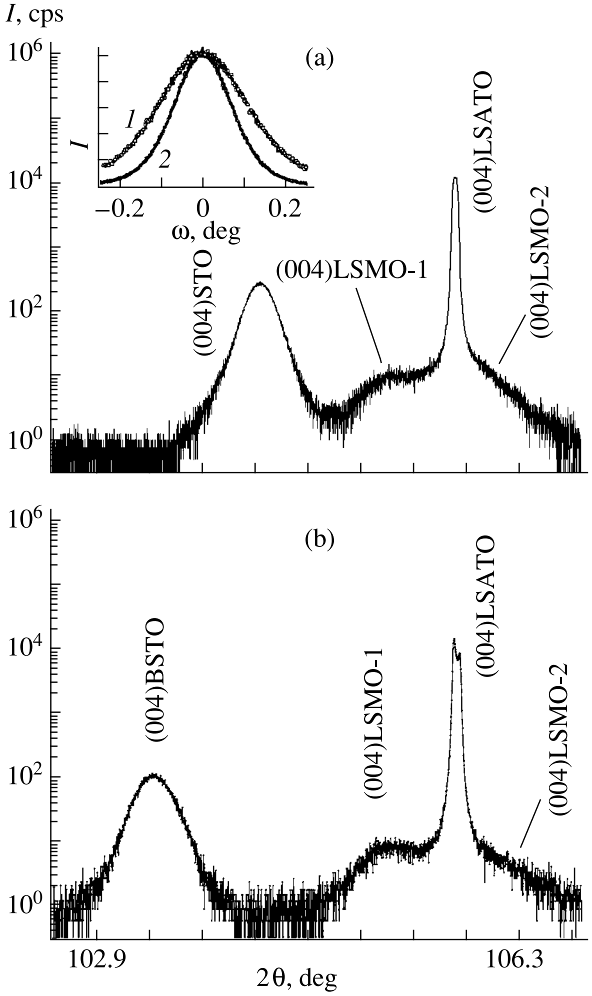


Fig. 1. (a) Fragment of the X-ray diffraction pattern ($CuK_{\alpha 1}$ radiation, ω/2θ scan) for the (001)LSMO/(001)STO/(001)LSMO/(001)LSATO epitaxial heterostructure in the case when the incident and reflected beams lie in the plane normal to the substrate plane. The inset shows the ω/2θ rocking curves for the (002) reflection from the intermediate (*1*) BSTO and (*2*) STO layers in the LSMO/BSTO/LSMO/LSATO and LSMO/STO/LSMO/LSATO heterostructures. (b) Fragment of the X-ray diffraction pattern for the (001)LSMO/(001)BSTO/(001)LSMO/(001)LSATO epitaxial heterostructure that is taken under identical conditions.

Capacitance C and dielectric loss tangent $\tan\delta$ of the parallel-plate film capacitors were measured with a Hewlett–Packard LCR 4263A meter (f = 1–100 kHz) with and without bias voltage $V_b = \pm 2.5$ V applied to the manganite electrodes. Permittivity ε of the intermediate insulator was calculated by the formula $\varepsilon = Cd/S$. The bias voltage was assumed to be positive when the upper electrode was positively biased. Resistance R of the manganite electrodes was measured by the van der Pauw technique with the same LCR meter (f = 100 Hz), and resistivity ρ was calculated by the formula $\rho = Rd_1\pi/\ln 2$ [8].

RESULTS AND DISCUSSION

Lattice mismatch $m = (a_s - a_f)/a_s$ (where a_s and a_f are the lattice parameters of the substrate and film, respectively) between the LSMO (pseudocubic lattice cell with a_1 = 3.871 Å [9]) and LSATO (pseudocubic lattice cell with a_2 = 3.868 Å [10]) is less than 0.1%. Therefore, a LSATO compound was chosen as a substrate for the LSMO/ferroelectric/LSMO heterostructures. The lattice parameter of $Ba_xSr_{1-x}TiO_3$ increases with x; accordingly, the lattice mismatch in the LSMO/BSTO system ($m \approx 1.5\%$) is nearly twice that in the LSMO/STO system ($m \approx 0.8\%$). The linear thermal expansion coefficients for LSMO, BSTO, and LSATO are close to each other [11, 12].

(i) Structure of the manganite and ferroelectric films. The X-ray patterns taken of the LSMO/STO/LSMO/(001)LSATO and LSMO/BSTO/LSMO/(001)LSATO heterostructures (Figs. 1a and 1b) exhibit peaks only from the substrate, intermediate ferroelectric layer, and manganite electrodes. Thus, the three-layer heterostructures grown were free of secondary macroprecipitates. The ω/2θ and ϕ scans indicate that the ferroelectric and manganite films had a clear-cut orientation both in and normally to the plane of the substrate. Below, the lower and upper manganite layers in the manganite/ferroelectric/manganite heterostructures will be denoted as LSMO-1 and LSMO-2, respectively.

The lattice parameters in the LSMO-1 layer of the LSMO/STO/LSMO heterostructure measured in the substrate plane, $a_{1p} = 3.870 \pm 0.005$ Å, and along the normal to the substrate plane, $a_{1n} = 3.878 \pm 0.005$ Å, were virtually coincident with those measured in the lower manganite layer of the LSMO/BSTO/LSMO heterostructure. The LSMO-1 layer in these heterostructures was structurally coherent to the substrate plane, which suggests that parameter a_{1p} equals the lattice parameter of the LSATO. The slight difference between a_{1p} and a_{1n} observed for the LSMO-1 may be due to biaxial compression stresses acting in the substrate plane. The effective volume of the lattice cell in the lower manganite electrode, $V_{eff} = a_{1p}^2 \times a_{1n} = 58.08$ Å^3, was close to the corresponding value for bulk stoichiometric LSMO samples (≈58.01 Å^3 [9]).

The lattice parameters of the 1000-nm-thick STO layer in the LSMO/STO/LSMO heterostructure measured in the substrate plane, $a_{\parallel} = 3.903 \pm 0.003$ Å, and normally to it, $a_{\perp} = 3.908 \pm 0.003$ Å, differed insignificantly from each other and from the lattice parameter for stoichiometric bulk STO samples. Nearly identical values (≈3.932 ± 0.003 Å) were also found for the parallel and normal lattice parameters of the 1000-nm-thick BSTO film in the LSMO/BSTO/LSMO heterostructure. At a thickness of the intermediate $Ba_xSr_{1-x}TiO_3$ film of 1000 nm, most of the mechanical stresses in it relax during growth and oxygen saturation.

The FWHM of the (002) reflection from the BSTO film in the LSMO/BSTO/LSMO heterostructure was measured to be 0.26°, which is roughly 50% larger than the FWHM of the (002) reflection from the STO in the LSMO/STO/LSMO heterostructure. The FWHM values for the STO and BSTO films in the heterostructures grown agree well with those for epitaxial ferroelectrics grown on oxide electrodes [11]. The degradation of the BSTO film microstructure in the heterostructures (compared with the microstructure of the STO film) may be attributed to the noticeably higher LSMO/BSTO lattice mismatch.

Because of the difference between the lattice parameters of LSMO and BSTO (STO), the LSMO-2 layer grown on the STO/LSMO/LSATO or BSTO/LSMO/LSATO was under tensile stresses acting in the plane of the substrate. The lattice parameter of the LSMO-2 in the STO/LSMO/LSATO and BSTO/LSMO/LSATO heterostructures measured normally to the substrate plane varied in the interval 3.85 and 3.86 Å, i.e., was markedly smaller than a_1. Since ($n0n$) reflections from the LSMO-2 and intense peaks from the substrate overlapped, the lattice parameter of the LSMO-2 in the plane of the substrate was difficult to measure.

(ii) Resistivity of the film LSMO electrodes. The temperature dependence of the resistivity of the 120-nm-thick LSMO film grown on the (001)LSATO is shown in Fig. 2. Resistivity ρ of this film increases with temperature in the range 77–400 K, the increase being the most pronounced between 330 and 360 K. Such behavior is consistent with resistivity data for bulk stoichiometric LSMO samples [5]. The resistivity of the manganite film remained almost the same after its surface had been covered by the 1000-nm-thick BSTO (STO) ferroelectric. (To measure the resistivity of the LSMO film in the BSTO/LSMO/LSATO and STO/LSMO/LSATO heterostructures, four silver contacts capacitively coupled with the manganite film were applied on the insulating layer.)

Starting with roughly the same temperature, the resistivity of the LSMO films grown on the BSTO/LSMO/(001)LSATO, STO/LSMO/(001)LSATO, and (001)LSATO drastically drops (Fig. 2). This means that the temperatures of spin ferromagnetic ordering in the LSMO-1 and LSMO-2 electrodes are almost the same. The resistivity of the LSMO-2 electrodes in the LSMO/BSTO/LSMO and LSMO/ STO/LSMO heterostructures exceeds that of the LSMO-1 electrodes by a factor of two to five (Fig. 2). To an extent, this is associated with grain boundaries present in the LSMO films (see inset in Fig. 2). As follows from the inset, grains in the LSMO-2 films were about 30 nm across on average. The resistivity of intergranular spacings in the manganite films may considerably exceed ρ inside the grains because of a high density of structural defects (primarily oxygen vacancies) and stoichiometry breaking [13].

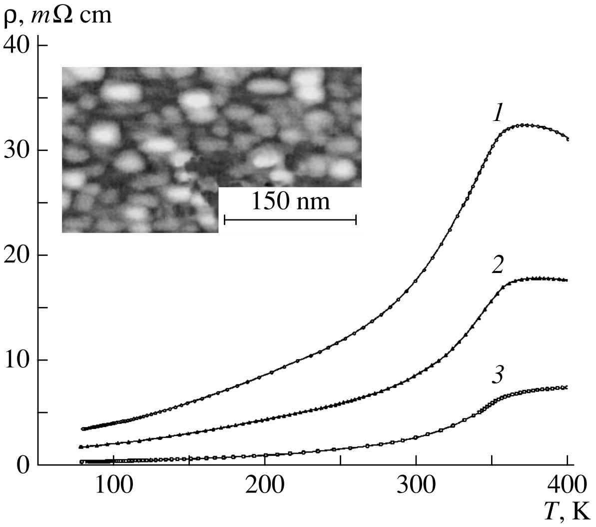


Fig. 2. Temperature dependences of resistivity ρ for the 120-nm-thick LSMO films grown on the (*1*) (001)BSTO/(001)LSMO/(001)LSATO, (*2*) (001)STO/(001)LSMO/(001)LSATO, and (*3*) (001)LSATO. The inset shows the AFM image of the free surface of the 120-nm-thick LSMO film grown on (001)STO/(001)LSMO/(001)LSATO.

(iii) Permittivity of the STO and BSTO films. The temperature dependences of permittivity ε for the 1000-nm-thick STO and 1000-nm-thick BSTO films in the LSMO/BSTO/LSMO and LSMO/STO/LSMO heterostructures are shown in Fig. 3. The permittivity maxima for the STO ($\approx 1300\varepsilon_0$) and BSTO ($\approx 1350\varepsilon_0$) are achieved at temperatures $T_m \approx 20$ and 100 K, respectively. Figure 3 also shows the temperature dependence of ratio $\varepsilon_0/\varepsilon$ for the STO and BSTO films at $f = 100$ kHz and $V_b = 0$. In a wide (≈150 K) temperature range, the curves $(\varepsilon_0/\varepsilon)(T)$ for the intermediate insulating layer in the LSMO/BSTO/LSMO and LSMO/STO/LSMO heterostructures obey the relationship

$$\varepsilon_0/\varepsilon = C_0^{-1}(T - T_{\mathrm{CW}}) + \varepsilon_0/\varepsilon_1, \qquad (1)$$

where parameter T_{CW} coincides with the Curie–Weiss temperature for STO (≈30 K [14]) and BSTO (≈145 K [15]) single crystals and coefficient C_0 equals $\approx 0.9 \times 10^5$ and $\approx 0.7 \times 10^5$ K for the STO and BSTO films, respectively (ε_0 is the permittivity of vacuum).

The values of C_0 obtained from the slopes (dashed lines) of the experimental curves $(\varepsilon_0/\varepsilon)(T)$ for the STO and BSTO films are in good agreement with the Curie constants for corresponding bulk crystals [15]. In Fig. 3, the values of $\varepsilon_0/\varepsilon_1$ for both types of the insulating films are shown by double-headed arrows. The value of ε_1 for the LSMO/STO/LSMO and LSMO/BSTO/LSMO heterostructures is temperature-independent and equals $2200\varepsilon_0$ and $2300\varepsilon_0$, respectively.

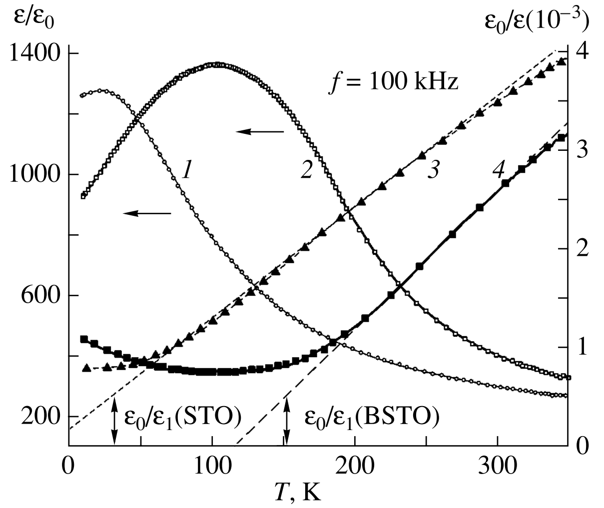


Fig. 3. Temperature dependences of permittivity (*1, 2*) $\varepsilon/\varepsilon_0$ and (*3, 4*) $\varepsilon_0/\varepsilon$ for the 1000-nm-thick (*1, 3*) STO and (*2, 4*) BSTO films in the parallel-plate capacitor structures with LSMO electrodes. The values of $\varepsilon_0/\varepsilon$ for the films are indicated by the double-headed arrows. The dashed lines are tangents to the curves $(\varepsilon_0/\varepsilon)(T)$ at temperatures for which $\varepsilon \sim (T - T_{CW})$.

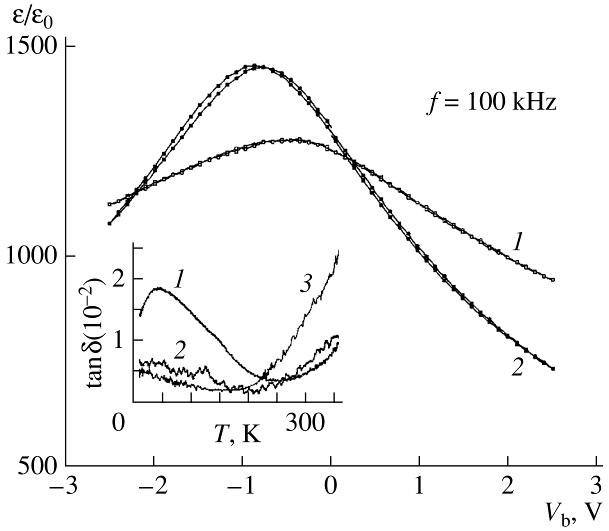


Fig. 4. Permittivity of the 1000-nm-thick (*1*) STO and (*2*) BSTO films in the LSMO/STO/LSMO and LSMO/BSTO/LSMO heterostructures vs. the bias voltage applied to the LSMO electrodes at (*1*) 10 and (*2*) 135 K. The inset shows the temperature dependences of tanδ for the 1000-nm-thick (*1*) BSTO and (*2, 3*) STO films in the above capacitor structures. The curves are taken at a frequency of (*1, 2*) 100 and (*3*) 10 kHz.

The first term on the right of (1) can be naturally assigned to the contribution of the intermediate insulator to the effective capacitance of the heterostructure; the second, to the contribution from the interfaces. The second contribution is due to electric field penetration into the manganite electrode [16]. The interface capacitance per unit area, $C_{\text{int}} = 2\varepsilon_1/d$, calculated for the STO/LSMO and BSTO/LSMO interfaces were found to be 3.9 and 4.1 μF/cm^2, respectively. These values are 30–50% lower than the corresponding capacitance induced by electric field penetration into a noble metal (Ag and Au) [16]. The reasons for the relatively low capacitance of the ferroelectric–manganite interface were considered in [4].

At $T = 250$–350 K, the values of tanδ ($f = 100$ kHz) for the 1000-nm-thick STO and BSTO films almost coincide, being equal to 5×10^{-3}–1×10^{-2} (see inset in Fig. 4). Note that, in this temperature range, both the STO and BSTO films are in the paraelectric phase. The growth of tanδ with temperature for both types of films is explained by an increase in their conductivity due to Frenkel–Pool emission [17]. As the frequency diminishes, the conductivity-related contribution to the loss tangent of the STO and BSTO films becomes prevailing (see inset in Fig. 4).

At $T < 200$ K, the loss tangent of the BSTO films far exceeds that of the STO films (see inset in Fig. 4). The reason for the different values of tanδ in these films is that, at $T < 145$ K, most of the thin BSTO film consists of ferroelectric domains and spontaneous polarization in the strontium titanate layers is absent at low temperatures. Domain wall relaxation significantly contributes to the dielectric losses of the ferroelectric phase. In a biaxially stressed epitaxial BSTO film, ferroelectric domains may form at temperatures several tens of degrees higher than the temperature of the ferroelectric phase transition in corresponding bulk single crystals. For this reason, the increase in tanδ of the BSTO films with decreasing temperature becomes appreciable even at $T \approx 200$ K.

The curves $\varepsilon(V_b)$ for the ferroelectric films in the LSMO/STO/LSMO and LSMO/BSTO/LSMO heterostructures exhibit a distinct maximum (Fig. 4). It is observed at negative bias voltages, which indicates the presence of an internal electric field in the STO and BSTO films, which is directed from the upper to lower electrode. This field is most fully compensated at $V_b = -0.5\ldots-0.8$ V. An internal electric field in these films emerges because the upper and lower interfaces in the three-layer film systems grown differ in microstructure and electrophysical parameters. At temperatures close to T_m, the relative change in the permittivity of the BSTO film, $[\varepsilon(V_b = 0) - \varepsilon(V_b = 2.5)]/\varepsilon(V_b = 0)$, is about 45%; for the STO film, this change is about 25%. At $T < 150$ K, the dielectric loss tangent decreases with rising positive bias voltage.

Note in conclusion that LSMO manganite electrodes favor the formation of a three-layer capacitor structure with a good epitaxial STO or BSTO layer between them. In LSMO/STO/LSMO and LSMO/BSTO/LSMO heterostructures, the capacitance induced by electric field penetration into the manganite electrode is roughly 4 μF/cm^2. This capacitance may have a considerable effect on experimental curves

$\varepsilon(T, V_b)$ taken of the intermediate ferroelectric in these structures. As the intermediate ferroelectric gets thinner, the effect of C_{int} on its effective permittivity becomes more pronounced.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research, grant no. 04-02-16738.

REFERENCES

1. M. Tachiki, M. Noda, K. Yamada, and T. Kobayashi, J. Appl. Phys. **83**, 5351 (1998).
2. Yu. A. Boikov and T. Claeson, Fiz. Tverd. Tela (St. Petersburg) **46**, 1231 (2004) [Phys. Solid State **46**, 1270 (2004)].
3. J. P. Hong and J. S. Lee, Appl. Phys. Lett. **68**, 3034 (1996).
4. Yu. A. Boikov and T. Claeson, Phys. Rev. B (in press).
5. G. J. Snyder, R. Hiskes, S. DiCarolis, *et al.*, Phys. Rev. B **53**, 14434 (1996).
6. Yu. A. Boikov and T. Claeson, Appl. Phys. Lett. **80**, 4603 (2002).
7. Yu. A. Boikov and V. A. Danilov, Pis'ma Zh. Tekh. Fiz. **30** (9), 26 (2004) [Tech. Phys. Lett. **30**, 361 (2004)].
8. T. I. Kamins, J. Appl. Phys. **42**, 4357 (1971).
9. B. Wiedenhorst, C. Hofener, Y. Lu, *et al.*, Appl. Phys. Lett. **74**, 3636 (1999).
10. *Crystec Specification.*
11. Yu. A. Boikov and T. Claeson, Physica B **311**, 250 (2002).
12. E. O. Wollan and W. C. Koehler, Phys. Rev. **100**, 545 (1955).
13. Yu. A. Boikov, D. Erts, and T. Claeson, Mater. Sci. Eng., B **79**, 133 (2001).
14. R. C. Neville, B. Hoeneisen, and C. A. Mead, J. Appl. Phys. **43**, 2124 (1972).
15. A. D. Hilton and B. W. Ricketts, J. Phys. D **29**, 1321 (1996).
16. H. Y. Ku and F. G. Ullman, J. Appl. Phys. **35**, 265 (1964).
17. Yu. A. Boikov and T. Claeson, Supercond. Sci. Technol. **12**, 654 (1999).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 896–903. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 80–87.
Original Russian Text Copyright © 2005 by Grekhov, Mnatsakanov, Yurkov, Tandoev, Kostina.

SOLID-STATE ELECTRONICS

Static and Dynamic Characteristics of an MOS-Controlled High-Power Integrated Thyristor

I. V. Grekhov*, T. T. Mnatsakanov, S. N. Yurkov**, A. G. Tandoev**, and L. S. Kostina***

** Ioffe Physicotechnical Institute, Russian Academy of Sciences,*
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
e-mail: konst@mail.ioffer.ru
*** Lenin All-Russia Institute of Electrical Engineering, Moscow, 111250 Russia*

Received November 11, 2004

Abstract—A theoretical analysis of physical processes in an MOS-controlled high-power integrated thyristor is presented. The design of this device and the effect of diffusion layer parameters on the *I–V* characteristic in the on state are considered. A rigorous calculation and estimates of the maximal anode current that can be turned off, which depends on the holding current of the thyristor structure shunted by an external MOS transistor, are made. This current is calculated as a function of the effective resistance, which includes the resistance of the MOS transistor channel and that of gate metallization. Simulation of the current decay shows that, as the MOS transistor is switched on, the current, after a delay, sharply (within several fractions of a microsecond) drops by about 90% and then goes on decreasing more smoothly. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

The insulated-gate bipolar transistor (IGBT) is today the basic device of high-voltage semiconductor electronics, which is able of handling powers in the range from several hundreds of watts to several hundreds of kilowatts. An IGBT chip is, in essence, a high-power circuit integrating 10^5–10^6 elements (cells) about 20 μm in size operating in parallel. Each cell is a high-voltage bipolar transistor the control circuit of which incorporates a low-voltage high-speed metal–oxide–semiconductor (MOS) field-effect transistor (MOSFET). The basic advantages of the IGBT over other devices in the power range mentioned above are extremely low switching energy losses in the control circuit and a high speed. However, the voltage of the IGBT in the on state is roughly twice that of a normal thyristor, all other things being equal. Therefore, efforts to design high-power field-controlled thyristor-based integrated devices (MOS-controlled thyristors (MCTs)) [4, 5] were begun almost immediately after the IGBT pioneering developments [1–3].

The equivalent circuit of an elementary cell in the MCT's simplest version is demonstrated in Fig. 1a. The thyristor (normally represented as consisting of n^+pn and pnp^+ transistors) is switched off by means of a p-channel MOSFET (p-MOSFET) connected in parallel to the n^+p emitter and integrated into the thyristor chip. Its p-channel, being formed by applying a negative bias to the transistor gate, short-circuits the n^+-emitter and p-base of the n^+pn transistor, thereby drastically decreasing the gain of the thyristor. The total gain of the n^+pn and pnp^+ transistors becomes less than unity, and the thyristor structure turns off. Figure 1b shows an MCT elementary cell with a p-channel MOSFET incorporated into the p-base. When the structure turns off, hole current J_h toward the p-base flows to the right under the n^+-region, then to the left through the p-channel under the gate, and finally through the auxiliary p^+-layer to the cathode. The total voltage drop on these regions is applied to the n^+p junction in the forward direction. Therefore, the current that can be switched off in the cell must generate a voltage lower than that at which the electron injection becomes appreciable (≈0.8 V for silicon). Thus, the smaller the total resistance of the circuit shunting the n^+p emitter, the

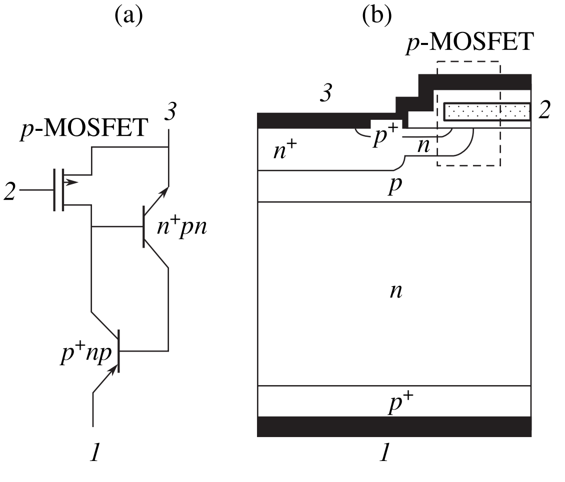


Fig. 1. Design of the MCT elementary cell: (*1*) anode, (*2*) gate, and (*3*) cathode.

higher the maximal anode current that can be turned off (hereafter, turned-off current) in the cell. This means, in particular, that the characteristic size of the cell should be small (no larger than 30 μm as a rule).

In spite of extensive investigations conducted by many companies, none of the MCT versions has been launched into mass production. This is primarily related to the complicated design of the elementary cell: for example, as many as 12 lithographic steps are necessary to fabricate the cell shown in Fig. 1b (versus eight such steps for the standard IGBT).

A new approach to tackling this problem has been recently reported [6–8]. Instead of an MOS transistor integrated into a microthyristor cell, it was suggested that a commercial low-voltage high-power MOS transistor with a very small channel resistance be used for simultaneously shunting all the emitter junctions of an integrated thyristor, which combines more than 10^5 ≈20-μm microthyristors on a chip. Such a separation of the bipolar and MOS technologies allows designers to drastically simplify the elementary cell of the high-voltage chip.

The cross-sectional view of the p^+npn^+ structure of such a device differs appreciably from the structure of a conventional high-power thyristor. The width and doping level of the wide *n*-base are certainly nearly the same as in a conventional device and correspond to voltages withstood by the collector junction of 3–5 kV; however, the base and emitter of the controlled n^+pn transistor are narrower by one order of magnitude and the doping level of the *p*-base is much higher. The gain of this transistor is very high ($\alpha \approx 0.95$), which makes it possible to reduce the gain of the pnp^+ transistor, i.e., to decrease the lifetime of holes in the *n*-base. A high speed of the n^+pn transistor provides a fast fall of the current early in switching off. The subsequent fall of the current down to zero takes a longer time but is still relatively short because of the small lifetime of holes in the *n*-base. Therefore, the switching-off time and switching losses in such a device are much smaller than, e.g., in a conventional gate turn-off thyristor (GTO). The device described above was called the super GTO (SGTO). Its basic advantages over the IGBT are (i) a lower on-state voltage, (ii) a simpler fabrication process, and (iii) a higher (roughly by a factor of two) operating current density; over the GTO, (i) a much lower switching-off power, (ii) a higher rate of switching on and off, and (iii) lower switching losses.

In this work, we analyze physical processes taking place in the SGTO in the on state and upon switching off. The whys and wherefores of the design of the device are presented, the effect of the base layer parameters on the I–V characteristic in the on state is studied, the limiting value of the turned-off current is calculated, and the dynamics of current fall upon switching off is considered. Based on the calculation results, basic requirements for the electrophysical properties of the semiconductor structure of the integrated thyristor and a high-power MOSFET used to switch the thyristor off are formulated.

PHYSICAL PROCESSES IN THE INTEGRATED THYRISTOR

(i) Design of the device. A silicon chip of the thyristor under study integrates a large number of ~30-μm microthyristors operating in parallel and is fabricated basically by planar technology. All the microthyristors share a *pn* collector junction, which must withstand (block) a high voltage (3–5 kV). Therefore, either guard rings or a lightly doped ring-shaped region of the *p*-base should be provided along the circumference of the *pn* junction to avoid surface breakdown. The layout of the rings and ring-shaped region, which extend the space charge width at the site the *pn* junction emerges on the surface, are well known. We assume that surface breakdown is prevented by the latter technique and the lightly doped ring-shaped region is created by deep (≈100 μm) prediffusion of aluminum with a surface concentration of 7×10^{16} cm^{-3} followed by grinding and polishing of the diffusion layer to a residual depth of ≈20 μm. Then, the rapid diffusions of boron and phosphorus to produce a thin-base n^+p^+pn transistor are carried out. Figures 2a and 2b show two possible dopant

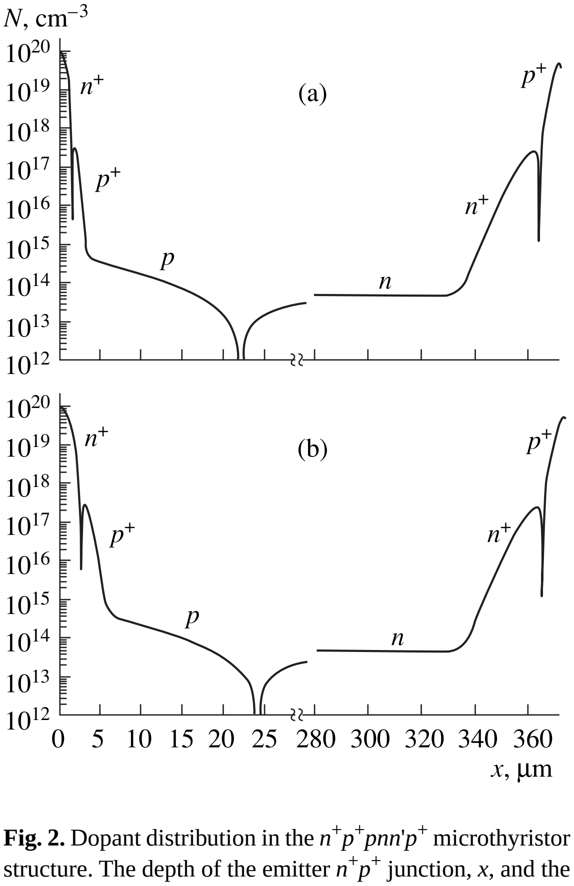


Fig. 2. Dopant distribution in the $n^+p^+pnn'p^+$ microthyristor structure. The depth of the emitter n^+p^+ junction, *x*, and the thickness of the p^+ layer, W_{p^+}, are, respectively, (a) 1.5 and 1.5 μm and (b) 2.5 and 2.5 μm.

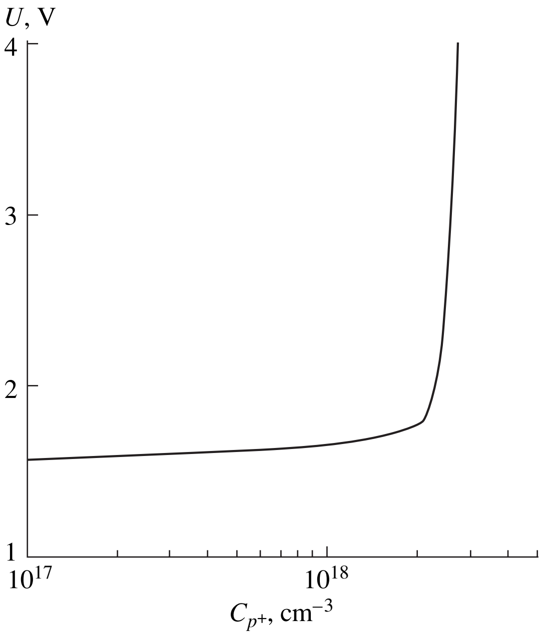


Fig. 3. On-state voltage drop U across the thyristor (the version shown in Fig. 2b) vs. the doping level in the p^+-layer at current density $j = 500$ A/cm^2.

profiles in the $n^+p^+pnn'p^+$ microthyristor structures, which differ in the depth of the n^+p^+ emitter junction and width of the heavily doped p^+-part of the p-base. The width of the lightly doped part (aluminum tail) of the p-base was taken such it was completely depleted at a voltage of ≈300 V. As the voltage rises further, the wide ring-shaped depleted region arising along the circumference of the common p^+-base of the chip prevents surface breakdown. In both structures, the width and resistivity of the n-base are 310 μm and $\rho \approx 100\ \Omega$ cm, respectively; the n'-layer is 32 μm wide, and the p^+-emitter is 8 μm wide. These values correspond to the standard unidirectional thyristor with a blocking voltage of ≈3 kV.

(ii) Effect of the doping level in the p^+ base on the *I–V* characteristic in the on state. MOS control of the thyristor will be effective if the shunt resistance is as low as possible. Most of this resistance is the spread resistance of the p^+-base under the n^+-emitter.

To decrease this resistance, it is necessary to raise the doping level of the p^+-layer and make it thicker. In this case, however, the transfer coefficient of the n^+p^+pn transistor decreases (i.e., the on-state voltage grows), especially at high forward current densities, since, as the doping level rises, both the injection coefficient of the n^+p^+-emitter and Schockley–Read electron lifetime τ_n, [9]

$$\tau_n = \frac{\tau_{n0}}{1 + (P/P_{\mathrm{pr}})}, \qquad (1)$$

decrease. In formula (1), P is the dopant concentration in the p^+ layer and P_{pr} is a constant depending on the process parameters and varying between 7×10^{15} and 1×10^{17} cm^{-3} [8].

The *I–V* characteristics were calculated for different doping levels in the p^+-base by numerically solving a fundamental set of equations, including the Poisson equation and continuity equations for electrons and holes.

The Issledovanie [10] quasi-one-dimensional application program used in the calculation takes into consideration the complete set of physical processes governing carrier transfer: recombination through deep centers, high-field effects, high-injection effects (electron–hole scattering, Auger recombination), and high-doping effects (band gap narrowing; a reduction of the mobility, diffusion coefficients, and Schockley–Read lifetime; and Auger recombination).

Figure 3 plots on-state voltage U in the thyristor structure shown in Fig. 2b against the doping level in the p^+-layer for current density $j = 500$ A/cm^2. It was assumed in the calculation that the carriers recombine through a level located at the midgap and the relationship $\tau_{p0} = 3\tau_{n0}$ typical of the silicon structures is valid (τ_{p0} is the hole lifetime in the wide lightly doped base of the thyristor). The value of P_{pr} was set equal to 1×10^{17} cm^{-3}, and τ_{p0} was taken to be ultimately high ($\tau_{p0} = 45$ μs). It is clear from Fig. 3 that, when the surface concentration of boron, N_{sB}, exceeds 2×10^{18} cm^{-3}, the sharp growth of U is observed. This is because the structure becomes unsaturated when the transfer coefficient of the n^+p^+pn transistor declines. In light of this, decreasing the spread resistance of the p^+-base by raising the doping level seems to be a great challenge.

Figure 4 plots U versus hole lifetime τ_{p0} in the n-base at lower current densities (100 and 200 A/cm^2). The curves are calculated for two values of P_{pr} for the two versions of the $n^+p^+pnn'p^+$ structure. From these curves, it becomes clear, in particular, how important application of advanced technologies (providing a high value of P_{pr}) is in fabricating the devices. For example, at a reasonable value of U, say, $U = 2$ V, an increase in P_{pr} from 7×10^{15} to 1×10^{17} cm^{-3} decreases τ_{p0} almost threefold, which substantially improves the dynamic characteristics, as will be shown below.

In general, the calculation demonstrates that the parameters of the diffusion layers of the n^+p^+pn transistor that are presented in Figs. 2a and 2b are readily reproducible by the standard bipolar technology and provide a good *I–V* characteristic of the integrated thyristor.

(iii) Peak turn-off capability. As was already noted, when the thyristor is switched off via shunting the n^+p^+-emitter with a MOSFET, the net resistance of the shunting circuit (comprising the spread resistance of the p^+-layer under the emitter, the MOSFET channel resistance, and the resistance of gate and emitter metallizations) plays a decisive part.

The design of the elementary cell (microthyristor) is shown in Fig. 5. Let us estimate the maximal density of turned-off current J_m for two cell layouts: strip (the rectangular cell is shunted on its opposite sides) and circular. The value of J_m for the square layout lies between these two estimates.

Let S_a be the active surface area of the device ($S_\mathrm{a} = S_1N$, where S_1 is the surface area of the microthyristor emitter and N is the number of microthyristors) and R_m be the MOSFET channel resistance plus the resistance of gate and emitter metallizations.

The thyristor is blocked by switching on the external MOSFET, which shunts the n^+p^+ emitter junction. It seems obvious that blocking will take place only if the holding current in the on state of the thyristor shunted is higher than the anode current passing through the device at the time of blocking. This criterion is taken to calculate J_m.

Let us find the holding current for the structure shown in Fig. 5. Clearly, just this value will be the maximal turned-off current. The calculations which follow are based on the approach worked out in [11, 12], which relies on the balance of mobile charge carriers in the p-base.

It is assumed that the injection into the n-base is high and into the p-base low, injection from the emitter p^+n' junction cross section is uniform, and the hole current in the n-base is close to one-dimensional. These assumptions are valid, since the thyristor cell size is much smaller than the thickness of the n-base.

The voltage distribution over the cathodic emitter junction is given by

$$\frac{d^2\vartheta}{dx^2} = a[J_{sn}(\exp\vartheta - 1)(1-\beta_n) - J_0], \qquad (2)$$

where $\vartheta = U_{n^+p}/\varphi_T$ is the normalized voltage across the emitter junction,

$$a = \frac{\rho_p}{W_p}\frac{1}{\varphi_T},$$

ρ_p/W_p is the sheet resistance of the p^+-base, β_n is the coefficient of electron transfer through the p^+-base, and J_0 is the density of the uniformly distributed hole current through the collector (central) junction.

The calculation is carried out in terms of the maximal turned-off current density, which is defined as $J_\mathrm{m} = I_\mathrm{m}/S_\mathrm{a}$, and hole current density J_0, which is related to J_m through the obvious relationship $J_0 = J_\mathrm{m}\alpha_p$, where α_p is the transfer coefficient of a wide-base transistor. Since recombination in the Narrow p^+-base is weak, we may put $\beta_n \approx 1$ and neglect the first term on the right of (2). Thus, the hole current is the major contributor to the distribution of $\vartheta(x)$. It enters the p-base with a constant density and then drains down into the gate. The voltage drop due to the passage of this current through the p^+-base under the emitter provides forward bias $\vartheta(x)$ of the emitter junction.

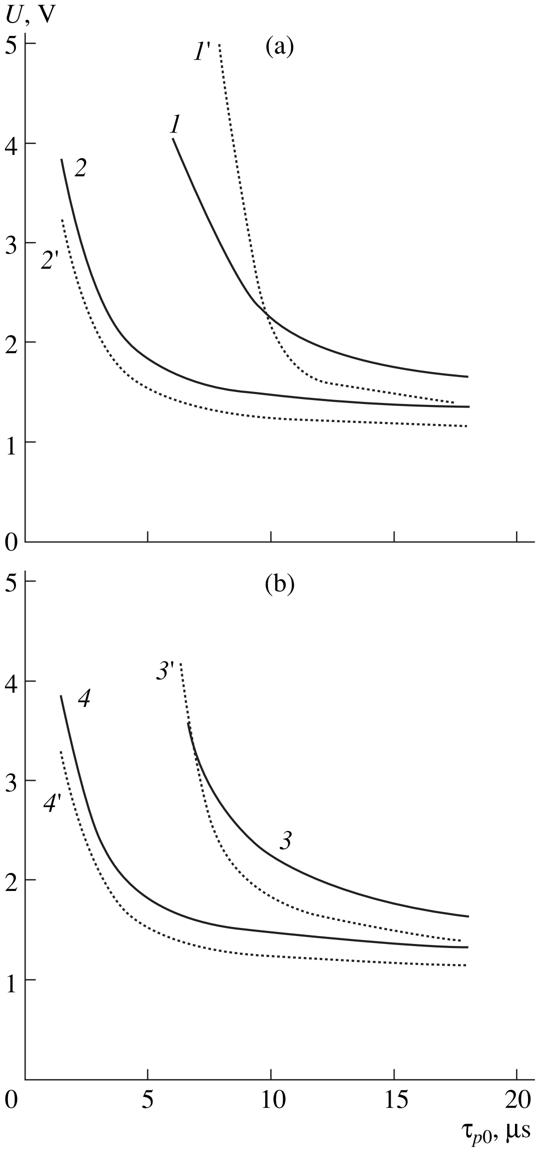


Fig. 4. On-state voltage drop U across the thyristor vs. the hole lifetime in the n-base at j = 100 (dotted lines) and 200 A/cm^2 (continuous lines) for the two versions of the thyristor design. (*1*, *1'*, *2*, *2'*) Version (a) in Fig. 2 and (*3*, *3'*, *4*, *4'*) version (b) in Fig. 2. P_pr = (*1*, *1'*, *3*, *3'*) 7×10^{15} and (*2*, *2'*, *4*, *4'*) 1×10^{17} cm^{-3}. The surface concentration of the dopant (boron) in the p^+-base is $N_\mathrm{sB} = 1 \times 10^{18}$ cm^{-3}.

The boundary condition at the left boundary follows from the symmetry of the structure,

$$\left.\frac{d\vartheta}{dx}\right|_0 = 0. \qquad (3)$$

At the right end of the structure, the gate potential relative to the cathode is specified,

$$\vartheta(x_\mathrm{g}) = U_\mathrm{g}/\varphi_T. \qquad (4)$$

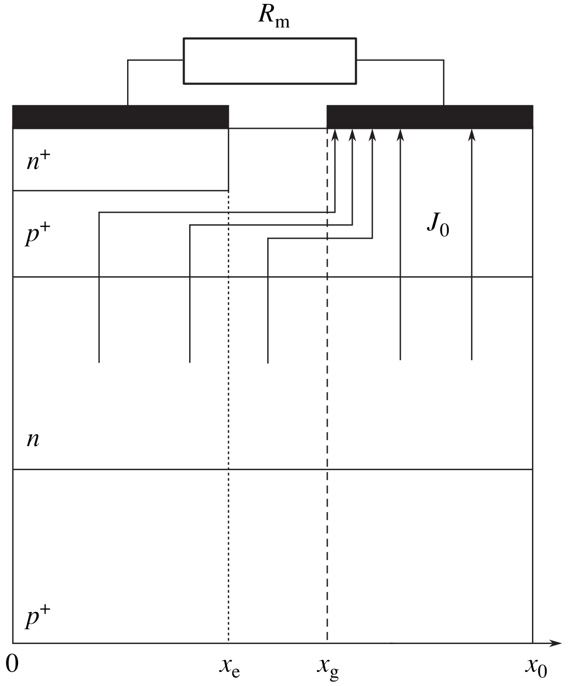


Fig. 5. Structure fragment clarifying the calculation of the microthyristor's basic characteristics.

Voltage U_g is related to the voltage drop across resistance R_m including the resistance of cathode metallization and the MOSFET channel resistance,

$$U_g = J_0 S_a R_m \quad (5)$$

(here, a small voltage drop across the portion $x_g - x_e$ is neglected).

With the above assumptions, Eq. (2) with boundary conditions (3) and (4) takes the form

$$\vartheta(x) = aJ_0\frac{1}{2}(x_e^2 - x^2) + \frac{U_g}{\varphi_T}, \quad (6)$$

and the maximal potential at the cathodic junction is given by

$$\vartheta_0 \equiv \vartheta(0) = aJ_0\frac{1}{2}x_e^2 + J_0 S_a R_m/\varphi_T. \quad (7)$$

Hence,

$$\vartheta(x) = \vartheta_0 - aJ_0\frac{x^2}{2}. \quad (8)$$

As was mentioned earlier, quantities J_0 and J_m are related as

$$J_0 = J_m\alpha_p. \quad (9)$$

At the holding point,

$$I_n\beta_n + I_m\alpha_p = I_m. \quad (10)$$

Here,

$$I_n = NL\int_0^{x_e} J_{sn}\exp\vartheta(x)dx = NLJ_{sn}\exp(\vartheta_0)\sqrt{\frac{\pi}{2aJ_0}}\operatorname{erf}\left(x_e\sqrt{\frac{aJ_0}{2}}\right), \quad (11)$$

is the current of the electrons injected by the cathodic emitter junction, L is the cell length, and N is the number of cells. Eventually, we have

$$I_n = I_m\frac{(1-\alpha_p)}{\beta_n} = NLJ_{sn}\exp(\vartheta_0)\sqrt{\frac{\pi}{2aJ_0}}\operatorname{erf}\left(x_e\sqrt{\frac{aJ_0}{2}}\right). \quad (12)$$

From (12), one can find the maximal potential across the emitter n^+p junction in the holding state by expressing it through holding current density $J_m = I_m/(x_0LN) = I_m/S_a$,

$$\vartheta_0 = \ln\left[\frac{J_m}{J_{sn}}\left(\frac{(1-\alpha_p)}{\beta_n}\right)x_0\sqrt{\frac{2a\alpha_p J_m}{\pi}}\frac{1}{\operatorname{erf}(x_e\sqrt{a\alpha_p J_m/2})}\right]. \quad (13)$$

The holding current density is convenient to find from expression (7), which can be represented as follows:

$$\vartheta_0 = J_0\left[\frac{S_a R_m}{\varphi_T} + a\frac{x_e^2}{2}\right]. \quad (14)$$

Then in view of (9), we get

$$J_m = \frac{U_0}{\alpha_p\left[\frac{\rho_p}{W_p}\frac{x_e^2}{2} + S_a R_m\right]}, \quad (15)$$

where $U_0 = \vartheta_0/\varphi_T$.

From expressions (13) and (15), one can determine unknowns ϑ_0 and J_m.

Note that, for a silicon *pn* junction, we can put ϑ_0 = 0.8 V with a reasonable accuracy.

Similar calculations for a circular n^+-emitter of radius R yield

$$J_m = \frac{U_0}{\alpha_p\left[\frac{\rho_p}{W_p}\frac{R_e^2}{4} + S_a R_m\right]}. \quad (16)$$

Hole current $J_m\alpha_p$ picked up by the collector junction enters the p^+-base and almost wholly (since recombination is weak) passes through the base toward the gate and then through the shunting circuit. The voltage drop due to the passage of this current through the spread resistance of the p^+-base and external resistance R_m provides a forward bias of the cathodic emitter junction. Knowing the sheet resistance of the p^+-base, emitter size, and the resistances of the MOSFET channel and base metallization, one can find the maximal value of the direct turned-off current. Since we are interested in the current and not in its density, relationships (15) and (16) are convenient to represent in the form

$$I_m = \frac{U_0}{\alpha_p[R_{sp} + R_m]},$$

where R_{sp} is the spread resistance of the p^+-base.

For the strip layout, R_{sp} can be represented as (see expression (15))

$$(R_{sp})_{str} = \left(\frac{\rho_p}{W_p}\right)\frac{x_e^2}{2}\frac{1}{S_a};$$

for the circular layout,

$$(R_{sp})_c = \left(\frac{\rho_p}{W_p}\right)\frac{R_e^2}{4}\frac{1}{S_a}.$$

The calculation of sheet resistance $\rho_p W_p^{-1}$ of the p^+-base under the n^+-emitter carried out using a dedicated subroutine showed that $\rho_p W_p^{-1} = 2.12 \times 10^4$ Ω for the structure shown in Fig. 2a and 1.59×10^4 Ω for the structure in Fig. 2b.

Based on the relationships obtained, we estimated the maximal value of turned-off current, I_m, for a device consisting of 8000 microthyristors with 30 × 30-μm emitters with the dopant distributions shown in Figs. 2a and 2b. If the carrier lifetime in the n-base at a high injection level is $\tau_{hl} = 15$ μs (which corresponds to $\alpha_p \approx 0.35$) and the net resistance of the MOSFET channel and metallization is $R_m = 0.028$ Ω, I_m for the strip layout equals 12.2 (distribution shown in Fig. 2a) and 15.8 A (distribution in Fig. 2b); for the circular layout, I_m = 21.3 and 25.7 A for the distributions shown in Figs. 2a and 2b, respectively.

The value of R_m, $R_m = 0.028$ Ω, is to an extent tentative: it is generally desirable that this value be much lower than the spread resistance of the p^+-base in order not to adversely affect the blocking property of the device. Since the emitter–base voltage in the integrated thyristor does not exceed 20 V even in the case of forced switching, the MOSFET switching off the structure may be low-voltage. Today, high-power MOS transistors with the operating voltage range 30–60 V and an extremely low channel resistance are in commercial-scale production. For example, the Infinion SPP80N06S2L-05 transistor is designed for a mean current of 300 A and a voltage of 55 V and has a channel resistance of 0.0038 Ω. If it is assumed that, e.g., 25% of the critical voltage (0.8 V) in the shunting circuit drops across the channel resistance and the rest 75%, across the spread resistance and metallization resistance of the chip (consisting of 8000 microthyristors with an emitter size of 30 × 30 μm), such a transistor may switch off a current of ~50 A.

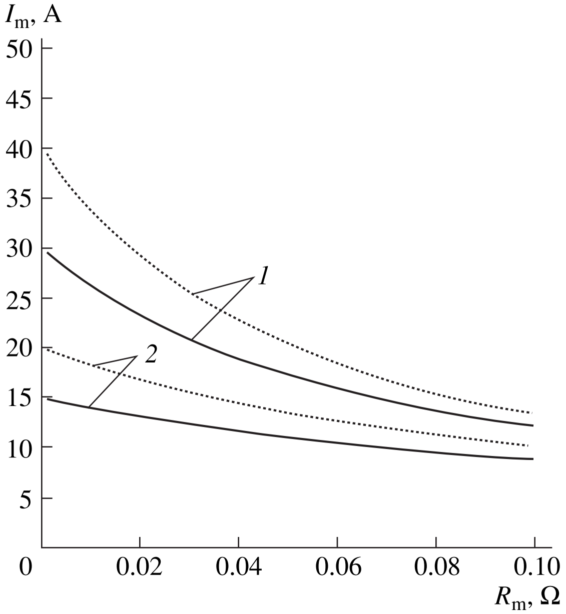


Fig. 6. Maximal turned-off current vs. resistance $R_m = R_{MOS} + R_{met}$, where R_{MOS} is the MOSFET channel resistance and R_{met} is the metallization resistance. The continuous curves, the version shown in Fig. 2a; the dotted curves, the version shown in Fig. 2b. (*1*) Circular and (*2*) strip layout.

Estimates show that, for the given layout, the metallization resistance turns out to be $R_{met} = 0.008$ Ω. Figure 6 plots maximal turned-off current I_m against resistance R_m, which includes the MOSFET channel and metallization resistances, $R_m = R_{MOS} + R_{met}$, for both design versions of the microthyristor (Figs. 2a and 2b). It is seen that the design shown in Fig. 2b combined with the circular layout far exceeds the version corresponding to Fig. 2a combined with the strip layout and that a decrease in R_m makes it possible to considerably raise I_m.

(iv) Dynamics of current decay upon switching off. The switching-off process was simulated with the Issledovanie application program [9]. Figure 7 shows the simulation results for the current switching-off transient at the initial current density of 200 A/cm^2 for the chip consisting of 8000 microthyristors (see the previous subsection) with $\tau_{hl} = 7$ μs. For the contributions from different components of the shunting circuit to be more illustrative, spread resistance R_{sp} and R_m were taken to be equal to 0 and 0.028 Ω, respectively (curve *1*); 0.075 and 0.028 Ω (curve *2*); and 0.1 and 0.028 Ω (curve *3*). Figure 8 shows the current switch-off transient for τ_{hl} varying from 7 to 20 μs at $R_{sp} = 0.075$ Ω and $R_{MOS} = 0.028$ Ω.

The results demonstrated in Figs. 7 and 8 indicate the presence of a delay early in the transient. Qualita-

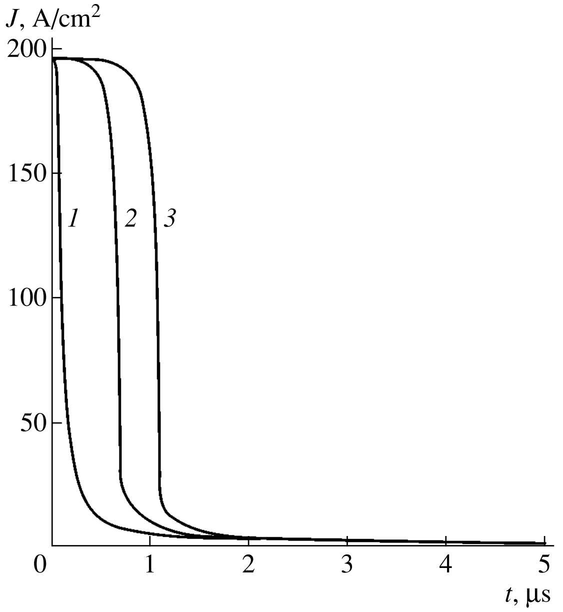


Fig. 7. Current switching-off transient in the thyristor chip. The spread resistance of the p^+-base R_{sp} is (*1*) 0, (*2*) 0.075, and (*3*) 0.1 Ω.

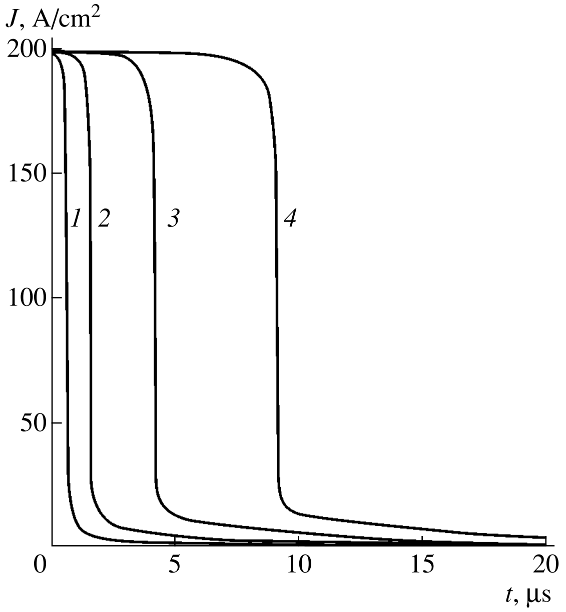


Fig. 8. Current switching-off transient in the thyristor chip. The minority carrier lifetime in the *n*-base under high injection τ_{hl} is (*1*) 7, (*2*) 10, (*3*) 15, and (*4*) 20 μs.

tively, they may be explained as follows. In the on state, the *p*- and *n*-bases of the thyristor structure are filled with an electron–hole plasma and the collector *pn* junction is forward-biased. Once the MOSFET is switched on, the electron injection from the n^+p^+ emitter decreases, since most of the hole current passes through the shunting circuit, the voltage drop across which is lower than the threshold of high electron injection (≈0.8 V). This causes the plasma to diffuse first in the *p*-base and then in the near-collector region of the *n*-base. The collector voltage changes sign, and the current sharply drops. The more slowly the plasma diffuses, i.e., the higher the shunt resistance and the higher the plasma concentration near the collector (the longer τ_{hl}), the longer the current decay delay is bound to be. These tendencies are distinctly seen in Figs. 7 and 8. At the same time, it should be noted that such a qualitative explanation does not take into account all features of the process as the turned-off current approaches a limiting value. This issue will be considered in a forthcoming study.

At low currents, the decay is heavily extended. This effect is akin to that observed in conventional GTOs and is associated with a large amount of the plasma near the anodic $p^+n'n$ injector, the amount of the plasma decreasing with decreasing τ_{hl}. Since the current drops slowly when the voltage across the device is already high, switching losses at this stage may be appreciable.

As was mentioned above, the basic factor that differentiates the device under consideration from the GTO is that the switching-off energy in the control circuit of the former (i.e., the energy necessary to charge the gate capacitance and form the control MOSFET channel) is very low. Therefore, the MOSFET may be on for most of the operating time and switched off immediately before applying an on pulse for the time of forward current passage. This allows designers to substantially raise the stability of the integrated thyristor against voltage spikes (the so-called *dU*/*dt* stability) in the off state and improve the temperature stability of the switching voltage.

CONCLUSIONS

The resistance of the circuit shunting the emitter n^+p^+ junction of the integrated thyristor in order to switch off the device is a basic parameter governing the switching-off dynamics and specifying the maximal value of the turned-off current. This resistance is the sum of the spread resistance of the p^+-bases of the microthyristors, metallization resistance of the thyristor chip, and resistance of the external MOSFET channel. The voltage drop across the shunting circuit must not exceed ≈0.8 V when the turned-off current passes through it. The numerical calculation of the forward-bias *I*–*V* characteristic with regard to nonlinear effects shows that an increase in the dopant concentration in the p^+-base of the microthyristor above 2×10^{18} cm^{-3} leads to a sharp increase in the forward voltage drop. Therefore, the width of the microthyristor's cathodic emitter should be shrunk to several tens of micrometers in order that the spread resistance be reduced to a reasonable level.

Analytical expressions relating the maximal turned-off current of the thyristor and the shunting circuit parameters are derived. The limiting turned-off current is estimated as a function of the spread resistance, chip metallization resistance, and MOSFET channel resistance. The numerical simulation of the current decay dynamics shows that, after the MOSFET has been switched on, the current, after a delay, sharply (within several fractions of a microsecond) drops by about 90% and then decreases more smoothly. In general, the calculation results demonstrate that the integrated thyristor switched off by a high-power low-voltage MOSFET offers good characteristics and may find wide application in medium- and high-power transducers. The fabrication process of the microthyristor chip appears to be rather simple; however, optimization of the microthyristor cell size needs special analysis.

REFERENCES

1. B. J. Baliga, *Power Semiconductor Devices* (PWS Publishing, Boston, 1994).
2. B. J. Baliga, M. S. Adler, R. P. Love, *et al.*, IEEE Trans. Electron Devices **31**, 821 (1984).
3. N. Thapar and B. J. Baliga, in *Proceedings of the 6th International Symposium on Power Semiconductor Devices and IC – ISPSD'94, Davos, 1994*, Abstract 4.3.
4. V. A. K. Temple, IEEE Trans. Electron Devices **33**, 1609 (1986).
5. F. Bauer, H. Hollenbeck, T. Stockmeier, and W. Fihtner, IEEE Electron Device Lett. **12**, 297 (1991).
6. Q. Huang, A. J. Amartunga, E. M. S. Narayaman, and W. I. Milne, IEEE Trans. Electron Devices **38**, 1612 (1991).
7. V. A. K. Temple, in *Proceedings of the European Pulsed Power Symposium, Saint Louis, 2002*, pp. 19/1–19/3.
8. Q. Huang, *CPES Report for Silicon Power Corporation* (Virginia Tech., 2001).
9. P. T. Landsberg and G. S. Kousik, J. Appl. Phys. **56**, 1696 (1984).
10. T. T. Mnatsakanov, I. I. Rostovtsev, and N. I. Philatov, Solid-State Electron. **30**, 579 (1987).
11. V. A. Kuz'min and S. N. Yurkov, Radiotekh. Élektron. (Moscow) **44**, 118 (1999).
12. V. A. Kuz'min, S. N. Yurkov, and A. G. Tandoev, Radiotekh. Élektron. (Moscow) **35**, 2389 (1990).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 904–913. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 88–96.
Original Russian Text Copyright © 2005 by Grekhov, Kyuregyan.

SOLID-STATE ELECTRONICS

Formation of the Space Charge Region in Diffusion *p–n* Junctions under High-Density Current Interruption

I. V. Grekhov* and A. S. Kyuregyan**

** Ioffe Physicotechnical Institute, Russian Academy of Sciences,*
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
*** Lenin All-Russia Institute of Electrical Engineering, Moscow, 111250 Russia*
e-mail: ask@vei.ru

Received December 2, 2004

Abstract—The recovery of diodes with diffusion *p–n* junctions in the case of high reverse current density j is analyzed. A condition for quasi-neutrality breaking in the diffusion layers with allowance for the dependence of charge carrier mobility μ on electric field strength E is obtained that is valid for a wide range of j. The problem of formation of the space charge region in a circuit with inductance L and resistance R is reduced to a system of two ordinary differential equations. Approximation of a numerical solution to this system makes it possible to derive crude analytical relationships between interrupted current density $\tilde{j}$, circuit parameters, diode parameters, and parameters of a forming voltage pulse (with amplitude V_m and pulse rise time t_p). The limiting parameters of a pulser with an inductive energy storage and current interrupter based on diffusion diodes are studied. The critical density of interrupted current $\tilde{j}_\mathrm{b}$ is determined at which the field in the space charge region near the anode reaches breakdown value E_b and intense impact ionization by holes begins. The impact ionization decreases the rates of current decay and voltage increase in the space charge region. As a result, at $\tilde{j} > \tilde{j}_\mathrm{b}$, t_p starts increasing and the overvoltage factor of the pulser decreases. The value of V_m corresponding to $\tilde{j} = \tilde{j}_\mathrm{b}$ is roughly given by $V_\mathrm{b} \approx m(\varepsilon \bar{v}_\mathrm{h} l_\mathrm{p}/\tilde{j}_\mathrm{b})^{1/2} E_\mathrm{b}^{3/2}$, where m is the number of diodes in the interrupter, ε is the permittivity of the semiconductor, $\bar{v}_\mathrm{h}$ is the saturated drift velocity of holes, and l_p is the depth of the *p–n* junction (diffusion depth). Theoretical predictions are confirmed by exact numerical simulation of the recovery process and qualitatively agree with the available experimental data.© *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Recovery of the blocking capacity of p^+–p–n–n^+ diodes with a deep diffusion *p–n* junction upon forward-to-reverse bias switching was first considered in [1, 2]. The analytical model of the process used in [1, 2], being qualitatively adequate, is inapplicable for quantitative description for two reasons. First, the case of constant reverse current is of minor practical interest. Second, the approximation of constant (independent of electric field E) carrier mobility μ restricts the domain of applicability of the theory to low reverse current densities and low voltages. Thus, for the most important (from the practical standpoint) case of high voltages and current densities in a circuit with inductance, an adequate theory of diode recovery upon direct-to-reverse bias switching is still lacking. The need for such a theory has greatly increased since the discovery of the effect of high-density (2–10 kA/cm^2) current fast interruption in diffusion silicon diodes [3], which was later called the SOS effect [4]. This effect has being successfully used for generation of ultra-high-voltage pulses with a rise time of 5–50 ns for a decade [5]. Numerical simulation [6–10] has demonstrated that the mechanism of this phenomenon is in complete qualitative agreement with the model put forward in [1, 2]. However, to gain deeper insight into the physics of the process and especially to find relationships between process characteristics, semiconductor device structure, and circuit parameters, an analytical or semi-analytical theory should be developed. In this work, we try to tackle this problem. To this end, the model used in [1, 2] is modified so as to take into account the real dependence $\mu(E)$ more rigorously. Specifically, at high current densities, the electroneutrality breaking condition, as well as the time and site of space charge formation in diffusion layers, is changed. Then, to describe the space charge region configuration, the constant drift velocity approximation is used, which is much simpler and more appropriate at high reverse voltages than the constant mobility approximation. Using this model to analyze interruption of the current passing through an inductive energy storage, we relate the circuit and diode parameters to the limiting amplitude and rise time of a voltage pulse across a resistive load.

1063-7842/05/5007-0904$26.00 © 2005 Pleiades Publishing, Inc.

1. STAGE OF HIGH REVERSE CONDUCTIVITY

Once a forward current pulse is off, the weakly doped central part of the diode is filled with a quasi-neutral electron–hole plasma, the distribution of which, $p(x) + N_d(x) = n(x) + N_a(x)$, along the x coordinate is a complicated function of the doping profile, $N(x) = N_d(x) - N_a(x)$; recombination characteristics of the diode; and current pulse waveform $J(t)$ (here, n, p, N_d, and N_a are the concentrations of electrons, holes, donors, and acceptors, respectively). An example of such a plasma distribution in a diode with a structure typical of a high-current-density interrupter [7] is illustrated in Fig. 1. Recovery of the blocking capacity of an interrupter consisting of m identical series-connected diodes takes place in the circuit shown in the inset to Fig. 1 [5–7]. Initially, current J in the circuit is zero, the capacitor is charged to voltage $-U_{c0}$, and the diode resistance is usually much lower than both load resistance R and internal circuit impedance $\sqrt{L/C}$. Therefore, at the first stage,

$$J(t) \approx -U_{c0}\sqrt{C/L}\sin(t/\sqrt{LC}). \tag{1}$$

This current pulls electrons and holes accumulated in the base mainly from the boundary regions, where abrupt concentration fronts that separate the plasma region from the regions free of nonequilibrium carriers move toward each other [1]. The velocity of the front moving from the cathode side, v_n^{f}, and that of the front moving from the anode side, v_p^{f}, are written, respectively, as [11]

$$\begin{aligned} v_n^{\mathrm{f}} &\equiv dx_n/dt = \Delta j_n(x_n)/q\Delta n(x_n) = \Delta j_p(x_n)/q\Delta p(x_n), \\ v_p^{\mathrm{f}} &\equiv dx_p/dt = \Delta j_n(x_p)/q\Delta n(x_p) = \Delta j_p(x_p)/q\Delta p(x_p), \end{aligned} \tag{2}$$

provided that the widths of the fronts, which are on the order of $D_{\mathrm{e,h}}/v_{n,p}^{\mathrm{f}}$ [12] are much smaller than all other characteristic dimensions of the problem (here, $x_{n,p}$ are the positions of the fronts, $D_{\mathrm{e,h}}$ are the diffusion coefficients, $\Delta j_{n,p}$ are the steps in the current densities, and Δn and Δp are the electron and hole concentrations at the fronts). At the initial stage of recovery, the dopant concentrations at the fronts, $|N(x_{n,p})|$, are large enough for the quasi-neutrality condition

$$\left|\frac{\partial E}{\partial x}\right| \ll \frac{q}{\varepsilon}|N(x_{n,p})| \tag{3}$$

to be also fulfilled beyond the plasma region (here, q is the elementary charge and ε is the permittivity of the semiconductor).

Since the condition $n + N = p$ is everywhere fulfilled and the concentrations and flows of minority carriers are negligible at $x < x_n$ and $x > x_p$, the velocities of the fronts obtained from (2) take the form

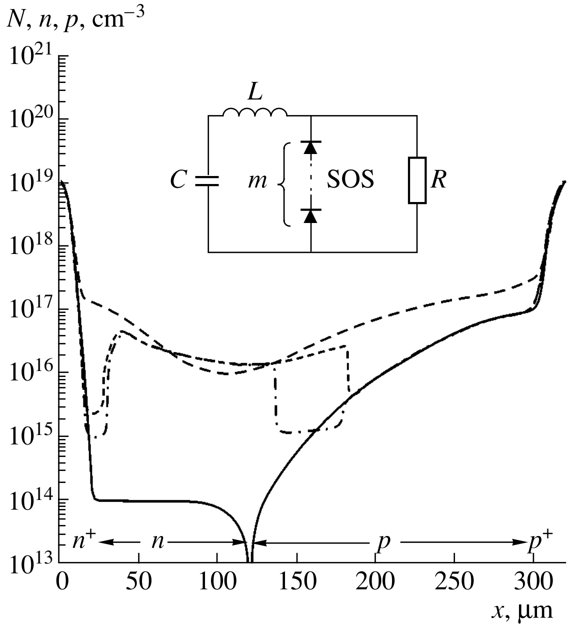


Fig. 1. Distributions of impurities (solid lines) and minority carriers in a diode of surface area 0.25 cm^2 at instant $t = 0$ of termination of the half-sinusoidal direct current pulse with a duration of 400 ns and amplitude of 450 A (dashed line), at instant $t = T_p$ of onset of anodic space charge formation (dotted line), and at the end of the pulse front $t = T_p + t_p$ (dash-and-dot line). The inset shows the circuit where current interruption takes place.

$$v_{n,p}^{\mathrm{f}} = \frac{j\mu_{\mathrm{e,h}}}{q(\mu_{\mathrm{e}}n_{n,p} + \mu_{\mathrm{h}}p_{n,p})}, \tag{4}$$

where $j = J/S$, S is the surface area of the diode, $\mu_{e,h}$ are the mobilities of electrons and holes, and $n_{n,p}$ and $p_{n,p}$ are the maximal concentrations of electrons and holes in the plasma region near boundaries $x_{n,p}$.

It is seen that $v_{n,p}^{\mathrm{f}}$ are merely the drift velocities of electrons and holes in the plasma concentration maxima at left and right fronts, respectively. As long as inequality (3) is fulfilled, the electric field strength outside the plasma is given by

$$j = qv_{\mathrm{e,h}}(E)|N(x)|. \tag{5}$$

In the simplest approximation of the dependences $v_{\mathrm{e,h}}(E)$, $v_{\mathrm{e,h}}(E) = \mu_{\mathrm{e,h}}E(1 + E/\bar{E}_{\mathrm{e,h}})^{-1}$ [11, 13], it follows from the Poisson equation and Eq. (5) that

$$|N(\tilde{x}_{n,p})| = \frac{1}{q}\left[\frac{j}{\bar{v}_{\mathrm{e,h}}} + \sqrt{\frac{\varepsilon j}{\gamma\mu_{\mathrm{e,h}}\lambda_{n,p}}}\right], \tag{6}$$

where $\bar{E}_{\mathrm{e,h}} = \bar{v}_{\mathrm{e,h}}/\mu_{\mathrm{e,h}}$, $\bar{v}_{\mathrm{e,h}}$ are the saturation velocities of electrons and holes, and $\lambda_{n,p}^{-1} = \partial\ln N(x_{n,p})/\partial x$. This formula determines positions $\tilde{x}_{n,p}$ of the fronts and concentrations $\tilde{N}_{n,\,p} = N(\tilde{x}_{n,\,p})$ such that the extent of electroneutrality breaking $\gamma \ll 1$ (i.e., $|\varepsilon\partial E/\partial x| \simeq \gamma q|N(x_{n,\,p})|$).

Hereafter, to simplify the mathematics, the real diffusion profiles will be approximated by simple exponentials,

$$N_{\mathrm{d,\,a}}(x) = \tilde{N}_{n,\,p}\exp\left(\frac{\tilde{x}_{n,\,p} - x}{\lambda_{n,\,p}}\right). \qquad (7)$$

It is easy to check that, e.g., for Gaussian profiles with surface concentrations $N^{\mathrm{s}}_{n,\,p}$ and diffusion depths $l_{n,\,p}$,

$$\lambda_{n,\,p} = \frac{l_{n,\,p}}{2}\left[\ln\left(\frac{N^{\mathrm{s}}_{n,\,p}}{N_0}\right)\ln\left(\frac{N^{\mathrm{s}}_{n,\,p}}{\tilde{N}_{n,\,p}}\right)\right]^{-1/2}, \qquad (8)$$

where N_0 is the donor concentration in the base.

At low current densities, the first term on the right of (6) is negligible and this formula coincides with the neutrality breaking condition obtained in [1]. However, at current densities typical of the SOS effect and at large diffusion depths, $j \gg \varepsilon v^2_{\mathrm{e,\,h}}/\mu_{\mathrm{e,\,h}}\lambda \approx 500$ A/cm^2; therefore, from (6), it follows that

$$\tilde{N}_{n,\,p} = N(\tilde{x}_{n,\,p}) \simeq j/q\bar{v}_{\mathrm{e,\,h}}. \qquad (9)$$

Equation (9) (or (6)) uniquely specifies $\tilde{x}_{n,\,p}$ only if the reverse current is constant. Otherwise (the most interesting case), the interruption current is unknown in advance. Yet, the problem basically can be solved if the dependences $j(t)$ and $p(x)$ are preset. Indeed, at high j, the displacement current, as well as the diffusion and recombination currents, is negligible within the plasma region [1]; therefore, the electron current density is

$$j_{\mathrm{e}} \simeq j\frac{\mu_{\mathrm{e}}}{\mu_{\mathrm{h}}}\left(1 + \frac{\mu_{\mathrm{e}}}{\mu_{\mathrm{h}}} + \frac{N}{n}\right)^{-1} \qquad (10)$$

and $\partial n/\partial t = -\partial j_{\mathrm{e}}/\partial x \simeq 0$ if the injection level is high ($n\mu_{\mathrm{e}} \gg N\mu_{\mathrm{h}}$). This means that concentration profile $p(x)$ remains "frozen" in the interval $x_n < x < x_p$ when the current changes sign [12]. Then, from (2) and (4), one can derive the "information-bearing" equations

$$\tilde{Q}_{n,\,p} \equiv \int_0^{T_{n,\,\mathrm{p}}} j(t)dt \simeq q\frac{\mu_{\mathrm{e}} + \mu_n}{\mu_{\mathrm{h,\,e}}}\left|\int_{x^0_{n,\,p}}^{x_{n,\,p}} p(x)dx\right|, \qquad (11)$$

which, along with (7) and (9), determine time instant $T_{n,\,\mathrm{p}}$ that the space charge starts to form, as well as the corresponding positions of the fronts, $\tilde{x}_{n,\,p}$, and current densities $\tilde{j}_{n,\,p}$. At $t = 0$, boundaries $x^0_{n,\,p}$ of the plasma-filled region are found from the equalities $N_{\mathrm{d}}(x^0_n) \simeq p(x^0_n) \equiv p^0_n$ and $N_{\mathrm{a}}(x^0_p) \simeq n(x^0_p) \equiv p^0_p$.

Before the formation of the space charge regions, the voltage drop across the diffusion layers exponentially grows, in accordance with an increase in their ohmic resistance, and, at $t = T_{n,\,\mathrm{p}}$, reaches the value

$$\tilde{U}_{n,\,p} \approx m\lambda_{n,\,p}\bar{E}_{\mathrm{e,\,h}}\ln\left(1 + \sqrt{\gamma\lambda_{n,\,p}\tilde{j}_{n,\,p}/\varepsilon v_{\mathrm{e,\,h}}\bar{E}_{\mathrm{e,\,h}}}\right), \qquad (12)$$

where subscripts n and p refer to the plasma-free regions near the cathode and anode, respectively.

The value of $\tilde{U}_n/m$ (usually not exceeding several volts) is far below voltage U_{c0} across the capacitor; therefore, in a first approximation, the time dependence of the current is described by formula (1) up to $t = T_n$. With this taken into account, simple estimates by formula (11) show that $\tilde{Q}_n < \tilde{Q}_p$ for the typical conditions under which the SOS effect is observed. That is, near the cathode, neutrality breaks earlier than near the anode, $T_n < T_{\mathrm{p}}$. This conclusion, reflecting the large diffusion depth l_p (and, thus, λ_p; see (8)) of acceptors, is supported by experiments [5] and numerical simulation.

2. FORMATION OF THE SPACE CHARGE REGION NEAR THE CATHODE

The cathodic space charge region, forming at $t > T_n$, expands rapidly until impact ionization due to an increase in the field strength starts in it. Considering that the drift velocities of carriers are saturated almost throughout the space charge region and assuming that $p_n \gg j/qv_n$ (the validity of this inequality means, in particular, that the displacement current in the space charge region can be neglected), one easily finds from (2) a relevant formula for the front velocity,

$$v^{\mathrm{f}}_n = \frac{j}{qp_n}\left[\xi\frac{\mu_{\mathrm{h}}}{\mu_{\mathrm{e}}} - \left(1 - \frac{1}{M_n}\right)\right], \qquad (13)$$

where $\xi = \mu_{\mathrm{e}}/(\mu_{\mathrm{e}} + \mu_{\mathrm{h}})$ and M_n is the coefficient of electron avalanche multiplication.[1]

Here, the static multiplication coefficient can be used, since the time constant of the establishment of the steady state in the space charge region (which exceeds transit time w_n/v_{e} by a factor of M_n) is far shorter than all other characteristic times of the problem. At $M_n = 1/\xi \approx 4/3$, the "cathodic" front might be expected to stop, because the intense outflow of holes from the plasma adjacent to the front (the process responsible for the motion and very existence of the front) is exactly compensated for by the inflow of holes generated in the space charge region by impact ionization. Actually, however, the front cannot stop completely. Indeed, an abrupt drop of v^{f}_n leads to an increase in front width

---

[1] Similar but not quite exact formulas were used previously in [8, 11].

D_e / v_n^f and depresses the diffusion of electrons from the plasma into the cathodic space charge. Therefore, the hole current in the space charge and, hence, multiplication coefficient M_n must be increased to maintain the ever growing total current.[2] However, in this case, the space charge density, which is proportional to $[(\bar{v}_h/\bar{v}_e + 1)M_n^{-1} - 1]$, decreases, which it is easy to check. As a result, M_n may increase only because of the expansion of the space charge region. Eventually, the front will move with a finite, though very low, velocity.

The value $M_n = 4/3$ in silicon diodes is attained when the voltage across the cathodic space charge region, $\hat{U}_n$, is approximately equal to half the "dynamic breakdown" voltage (see, e.g., [14]), which, in general, depends on $\tilde{j}_n$ and λ_n. For a shallow n^+–n junction, voltage $\hat{U}_n$ can be simply estimated under the assumption that $\hat{w}_n \gg \lambda_n$ (where $\hat{w}_n$ is the thickness of the cathodic space charge region) and that the charge distribution in this region is heavily asymmetric. One should only take into account that, in this case, the space charge density is equal to $\tilde{j}_n/\bar{v}_e$ almost throughout the cathodic space charge region and employ the well-known formulas [15] for the parameters of the space charge region in heavily asymmetric p–n junctions. Then, we get

$$\hat{U}_n \approx m\varepsilon\bar{v}_e E_b^2/4\bar{j}_n, \quad \hat{w}_n \approx \varepsilon\bar{v}_e E_b/\sqrt{2}\tilde{j}_n, \tag{14}$$

where E_b is the breakdown field, which will be considered to be a given parameter of the semiconductor for simplicity.

At large diffusion depths l_n, formulas (14) yield understated values of $\hat{U}_n$ and $\hat{w}_n$. More exact estimators are given in [16]. Time t_n of rapid expansion of the space charge layer and the time of voltage drop across it can be estimated in terms of order of magnitude if a multiplication-induced decrease in the front velocity is neglected,

$$t_n \approx \hat{w}_n / v_n^f(T_n) = q\varepsilon\bar{v}_e E_b \tilde{p}_n \mu_e / \sqrt{2}\varepsilon\mu_h \tilde{j}_n^2. \tag{15}$$

When deriving these formulas, we also assumed that the change in the current for a short period of time ($t_n \sim 10$ ns) is small (i.e., $j(t) \approx \tilde{j}_n$). For the above reason, the voltage across the space charge layer continues to grow at $t > T_n + t_n$, while much more slowly. The exponentially increasing resistance of the plasma-free part of the p-layer and the resistance of the central part of the plasma region itself, where the carrier concentration is minimal, make an additional contribution to the increase in the diode voltage at this time. By the time $t = T_p$, the total voltage drop is given by

$$U(T_p) \approx \tilde{U}_n + \hat{U}_n + \tilde{U}_p + U_{pn}, \tag{16}$$

where

$$U_{pn} = \tilde{j}_p \int_{x_n}^{x_p} \frac{dx}{qp(x)(\mu_n + \mu_p)}.$$

Finally, time $t = T_p$ manifests the onset of the SOS effect: the current through the diode is rapidly interrupted and the diode voltage starts growing because of the formation of the space charge region in the thick p-layer.

3. FORMATION OF THE SPACE CHARGE REGION IN THE DIFFUSION *P*-LAYER

Once the space charge starts forming in the plane $x = \tilde{x}_p$ of the diffusion p-layer, the velocity of the "anodic" plasma front is given by a formula similar to (13),

$$v_p^f = \frac{j}{qp_p}\left[\xi - \left(1 - \frac{1}{M_p}\right)\right]. \tag{17}$$

Since the coefficient of impact ionization for holes in Si is much smaller than that for electrons [15, 16, 18], the strong inequality $(M_p - 1) \ll 1$ is fulfilled until maximal field strength E_p^m in the anodic space charge region becomes nearly equal to breakdown field E_b [14]. This considerably simplifies analysis of the process, although two complicating factors appear.

First, the current through the diodes decreases considerably as the anodic space charge regions form. At the same time, the voltage drop across them, $U_p(t)$, increases sharply due to the presence of the inductance and can far exceed the voltage across the capacitance of the circuit. Therefore, the process can be described with a differential equation that takes into account the basic features of the circuit. In our case, this equation takes the form

$$\frac{L}{R}\frac{dU_p}{dt} + LS\frac{dj}{dt} + U_p = 0. \tag{18}$$

When deriving (18), we assumed that both the voltage drop across the high capacitance of the circuit and the voltage drop across the diode (except for the anodic space charge region) change weakly for pulse rise time t_p and are equal each other. The latter assumption is validated by the fact that the total current of the circuit reaches a maximum at $t = T_p$ in an appropriately designed pulser.

Second, one should take into account the nonuniform distribution of charged acceptors in the anodic space charge region (Fig. 2). In the saturated hole drift velocity approximation for diffusion profile (7), this

---

[2] As follows from estimates, the displacement current also contributes, though insignificantly, to the total current.

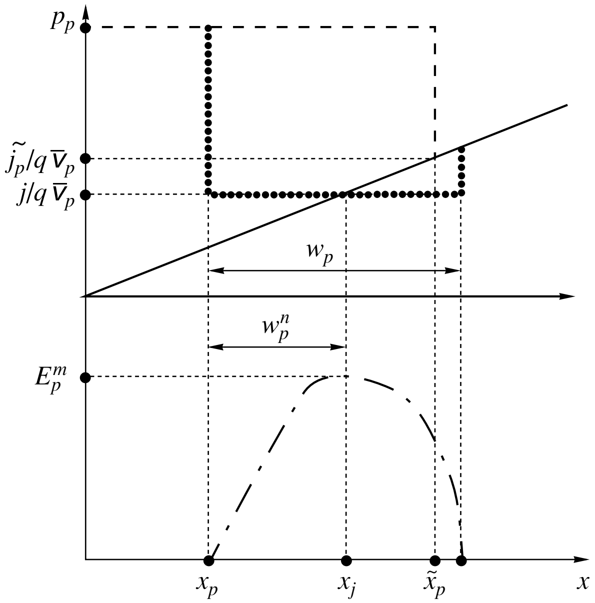


Fig. 2. Idealized distributions of acceptors (solid line), nonequilibrium carriers at $t = T_\mathrm{p}$ (dashed line) and $t > T_\mathrm{p}$ (dotted line), and field strength at $t > T_\mathrm{p}$ (dash-and-dot line) that were used in analysis of the space charge region formation in the p-layer.

can be done using relationships between w_p, U_p, E_p^m, and $w_p^n \equiv (x_j - x_p)$ given, e.g., in [16],

$$U_p = \frac{2m\lambda_p^2\tilde{j}_p}{\varepsilon\bar{v}_\mathrm{h}}u(y, \omega), \tag{19}$$

$$E_p^\mathrm{m} = \frac{\lambda_p\tilde{j}_p}{\varepsilon\bar{v}_\mathrm{h}}e(y, \omega), \tag{20}$$

$$w_p^n = -\lambda_p \ln\omega(\coth\omega - 1). \tag{21}$$

Here, $\omega = w_p/2\lambda_p$, $y = j/\tilde{j}_p$, $u(y, \omega) = yf(\omega)$, $e(y, \omega) = y\varphi(\omega)$, $f(\omega) = \omega(\omega\coth\omega - 1)$, and $\varphi(\omega) = [\omega(\coth\omega - 1) - 1 - \ln\omega(\coth\omega - 1)]$.

The coordinate of the dynamic p–n junction (i.e., of the plane where the field is maximal and the space charge density is zero) is calculated from (7) and (9),

$$x_j = \tilde{x}_p + \lambda_p \ln j/\tilde{j}_p. \tag{22}$$

The coordinate of the left-hand boundary of the space charge region (coincident with the coordinate of the right-hand front of the plasma region) is obtained by integration of (17),

$$x_p = \tilde{x}_p - \frac{\xi}{q\tilde{p}_p}\int_{T_\mathrm{p}}^{t} j\,dt, \tag{23}$$

under the assumption that the carrier concentration in the plasma at $t = 0$ is weakly dependent on x near $x = \tilde{x}_p$ (i.e., $p(x) \approx p(\tilde{x}_p) \equiv \tilde{p}_p$), as is observed in Fig. 2. Upon introduction of dimensionless parameters

$$\theta = \frac{\tilde{j}_p\xi}{q\tilde{p}_p\lambda_p}(t - T_\mathrm{p}), \quad \chi = \frac{\varepsilon\bar{v}_\mathrm{h}SR}{2\lambda_p^2 m}, \quad \zeta = \frac{q\tilde{p}_p\lambda R}{\xi\tilde{j}_p L}$$

and straightforward rearrangements, formulas (18)–(23) can be reduced to a system of two differential equations,

$$\frac{d\omega}{d\theta} = \frac{y[\chi + f(\omega)] - \zeta f(\omega)}{(\chi + \omega^2)[1 + f(\omega)/\omega]}, \tag{24}$$

$$\frac{dy}{d\theta} = -y\frac{y[\omega^2 - f(\omega)] + \zeta f(\omega)}{(\chi + \omega)^2}, \tag{25}$$

with obvious boundary conditions

$$\omega(0) = 0, \quad y(0) = 1. \tag{26}$$

The only exact result that follows from system (24)–(26) is that, at any χ and ζ, functions $y(\theta)$ decrease monotonically and functions $e(\theta)$, $u(\theta)$, and $w(\theta)$ reach maxima e_m, u_m, and ω_m, respectively, at time instants θ_E, θ_U, and θ_w such that $\theta_E < \theta_U < \theta_w$. In other words, first the field strength; then, the voltage drop across the space charge; and, finally, the space charge layer thickness reach maximal values. A numerical solution to system (24)–(26) was analyzed in a wide range of χ and ζ. It was found that the dependences $e_\mathrm{m}(\chi, \zeta)$, $u_\mathrm{m}(\chi, \zeta)$, and $\theta_{0.9}(\chi, \zeta)$ take the form

$$e_\mathrm{m}(x, \zeta) = \frac{1}{2}\frac{\chi^{3/5}}{1 + \zeta^{3/5}}F_E(\chi, \zeta), \tag{27}$$

$$u_\mathrm{m}(\chi, \zeta) = \chi^{9/10}(1 + 1.5\zeta^{9/20})^{-2}F_U(\chi, \zeta), \tag{28}$$

$$\theta_{0.9}(\chi, \zeta) = 2(\chi/\zeta)^{1/3}(1 + 0.1\chi^{-1/2})F_t(\chi, \zeta), \tag{29}$$

where $\theta_{0.9}$, which is determined from the condition $u(\theta_{0.9}) = 0.9u_\mathrm{m}$, characterizes rise time t_p of the voltage pulse across the load. Very weakly varying functions $F_{E, U, t}(\chi, \zeta)$ appearing in (27)–(29) differ from unity by less than 40% in the interval $0.01 < \chi, \zeta < 100$; therefore, we will put $F_{E, U, t}(\chi, \zeta) = 1$ for simplicity. Then, formulas (28) and (29) in the dimensional variables take the form

$$V_\mathrm{m} \approx \tilde{j}_p SR\left(\frac{2\lambda_p^2 m}{\varepsilon\bar{v}_\mathrm{h}SR}\right)^{1/10}\left[1 + \frac{3}{2}\left(\frac{qp_p\lambda_p R}{\xi\tilde{j}_p L}\right)^{9/20}\right]^{-2}, \tag{30}$$

$$t_p \approx 2\left(\frac{q\tilde{p}_p}{\xi\tilde{j}_p}\right)^{2/3}\left(\frac{\varepsilon\bar{v}_\mathrm{h}SR}{2m}\right)^{1/3}\left[1 + \frac{\lambda_p}{10}\left(\frac{2m}{\varepsilon\bar{v}_\mathrm{h}SR}\right)^{1/2}\right], \tag{31}$$

where V_m is the maximal voltage drop across the anodic space charge region.

These results certainly cannot be viewed as an exact quantitative description of the process.[3] However, they provide straightforward qualitative relationships between the parameters of the circuit, diode, and pulse being formed. As was expected, V_m increases and the rise time of the pulse, t_p, decreases as the current interrupted and load resistance grow, which is in full agreement with experimental results [5]. When R is very small (current generation mode), $V_m \sim \tilde{j}_p SR$; at high R (no-load mode), the dependences $V_m(R)$ and $t_p(R)$ saturate and the voltage pulse amplitude reaches a maximal value, $V_m \sim \xi \tilde{j}_p^2 SL/2q\tilde{p}_p\lambda_p$. The dependences of V_m and t_p on the other parameters are also quite clear. For example, the charge that must be extracted from the p-layer to extend the space charge region increases with the concentration of nonequilibrium holes. Therefore, all other factors being the same, the pulse amplitude is bound to decrease and the pulse rise time to increase with increasing $\tilde{p}_p$. In its turn, $\tilde{p}_p$ increases with the amplitude and duration of the direct pump current pulse. As diffusion depth l_p (and, hence, parameter λ_p) increases, the average space charge density decreases. Therefore, the voltage drop across the space charge region is bound to decrease with increasing λ_p, all other things being equal. Accordingly, the charge to be extracted for space charge formation, and, hence, the front duration, increase with λ_p, because v_p^f does not explicitly depend on l_p: it depends only on concentration steps Δn and Δp, as well as on current density steps $\Delta j_{n,p}$ at the front (see formula (2) and its consequences (13) and (17)). On the other hand, an increase in l_p must change the values of Δn, Δp, and $\Delta j_{n,p}$ (even though the other parameters of the circuit and diodes remain the same (see Section 5)) and, thereby, may indirectly influence the value of v_p^f. However, this issue is beyond the scope of the present study. It should also be noted that the dependences of V_m and t_p on the diffusion depth (formulas (30) and (31)) are the exact reverse of those obtained in [8] in the quasi-neutral approximation. At first glance, they also contradict the experimental data obtained in [5]. Actually, however, this contradiction is apparent and will be resolved below.

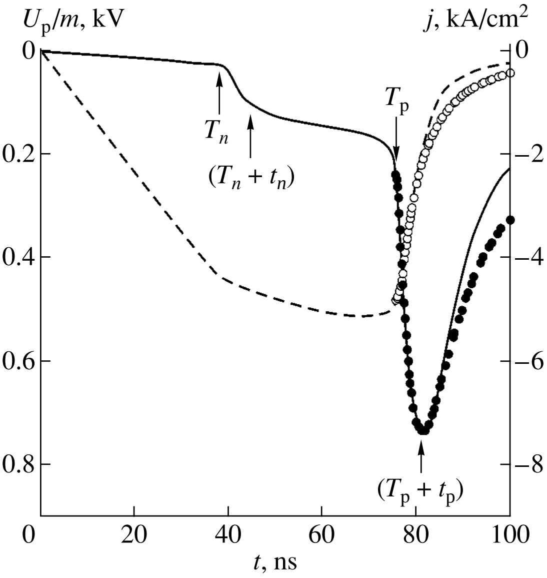


Fig. 3. Volt–second (solid line and filled circles) and ampere–second (dashed line and open circles) characteristics of the recovery of the current interrupter consisting of m silicon diffusion diodes (their structure is shown in Fig. 1). The characteristics are obtained by simulation of the recovery process with Issledovanie program [18] in the circuit with mc = 2 μF, L/m = 6.4 nH, R/m = 1.25 Ω, and U_{c0}/m = 200 V (lines) and by numerical solution of the system of equations (25) and (26) at $\chi = 0.43$ and $\zeta = 0.39$ (circles).

To check the conclusions of this section, we compared them with the results of exact numerical simulation of the recovery process in diffusion diodes using the Issledovanie program [17]. Both calculation methods turned out to coincide qualitatively and in many respects quantitatively. As an example, Fig. 3 demonstrates the recovery process for an interrupter consisting of m silicon diffusion diodes (the structure of the diodes and the nonequilibrium carrier initial distribution in them are shown in Fig. 1). In the calculations, the parameter values were the same as in [7]: S = 0.25 cm^2, λ_p = 18 μm (this value follows from (8) at l_p = 200 μm and $N_p^s = 10^{17}$ cm^{-3}), $\tilde{p}_p = 2.5 \times 10^{16}$ cm^{-3} (Fig. 1), $\tilde{j}_p$ = 4.75 kA/cm^2, L/m = 6.4 nH, and R/m = 1.25 Ω. In addition, we put mC = 2 μF and U_{c0}/m = 200 V in order to provide a desired value of the interrupted current, $\tilde{j}_p$, and a typical value of T_p ($T_p \approx$ 75 ns). The associated values of χ and ζ to be substituted into Eqs. (24) and (25) are $\chi = 0.43$ and $\zeta = 0.39$. In processing the results of numerical simulation, the equality $\max\rho(x) = 0.1qN_a(x)$ met on the right-hand side of the dynamic p–n junction ($\rho(x)$ is the space charge density) was taken to be the condition for neutrality breaking and space charge formation (in this case, $\gamma = 0.1$). At $t = T_p$, the voltage drop per diode, $U(T_p)/m$, reaches 235 V. Shifting the results of integration of system (24)–(25) along the vertical axis by T_p and along the abscissa axis by $U(T_p)/m$, one can compare the results obtained in terms of the simplified theory and those of the exact numerical simulation. As is seen from Fig. 3, agreement between the results is good up to the final phase of the process, when the applicability conditions for system (24)–(25) (namely, the constancy of the hole drift velocity, voltage drop across the cathodic space charge region, and hole concentration in the plasma region near $x = x_p$) are violated. The depen-

---

[3] To this end, system (24)–(25) must be solved numerically, which is today a trivial problem.

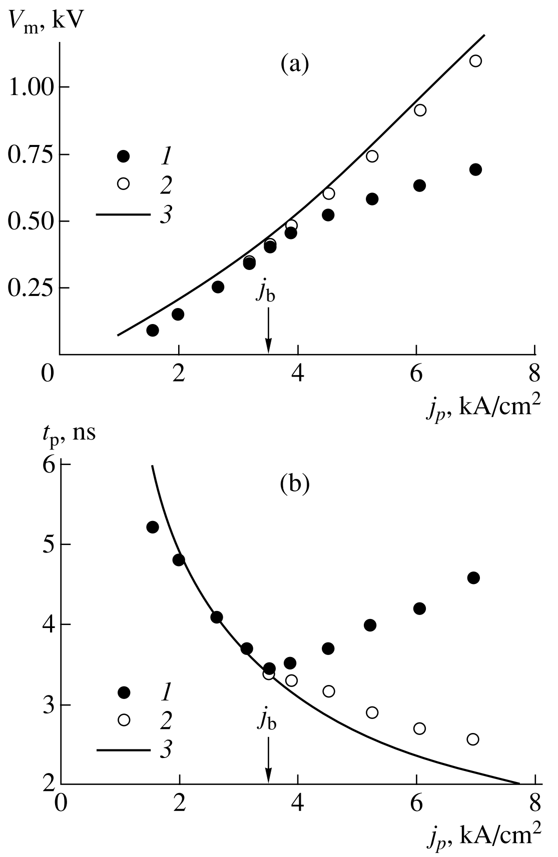


Fig. 4. Dependences of the (a) amplitude and (b) rise time of the voltage pulse in the anodic space charge region on the interrupted current density at $\tilde{p}_p = 2 \times 10^{16}$ cm^{-3} and R/m = 1.8 Ω. Numerical simulation (*1*) with and (*2*) without regard to impact ionization in the anodic space charge and (*3*) calculation by formulas (30) and (31). The arrows indicate the interrupted current density at which the maximal field strength in the space charge region reaches a breakdown value.

dences of V_m and t_p on the interrupted current density (Fig. 4) and load resistance (Fig. 5) that were calculated by formulas (30) and (31) also agree well with the results of numerical simulation when impact ionization in the anodic space charge region is disregarded.

4. MAXIMAL ATTAINABLE PARAMETERS OF CURRENT INTERRUPTERS

The maximal field strength in the space charge region, E_p^m, reaches breakdown value E_b when the interrupted current density rises up to a certain critical value $\tilde{j}_b$, which can be determined from formulas (20) and (27),

$$\tilde{j}_b \approx q\tilde{p}_p\lambda_p R/\xi L\zeta_b, \qquad (32)$$

where ζ_b is a solution to the equation

$$\zeta_b(1+\zeta_b^{3/5}) \approx \frac{q\tilde{p}_p R}{2\xi E_b L}\left(\frac{SR}{2m}\right)^{3/5}\left(\frac{\lambda_p^2}{\varepsilon \bar{v}_h}\right)^{2/5}. \qquad (33)$$

The value of V_m therewith reaches maximal value V_b, which is easy to find from the simple and physically clear formula

$$V_b \approx 3m(\varepsilon \bar{v}_h \lambda_p/\tilde{j}_b)^{1/2} E_b^{3/2}, \qquad (34)$$

which follows from (27) and (28) and has the same accuracy. Using the well-known relationships between the space charge parameters [15], it is easy to check that (34) relates the breakdown field to the breakdown voltage for a linear *p–n* junction with a space charge density gradient of $32\tilde{j}_p/81\bar{v}_p\lambda_p$. Such, indeed, must be the case in view of the fact that the breakdown voltages for linear and deep diffusion *p–n* junctions differ only slightly [16] and the current density at the instant the voltage reaches a maximum is approximately 81/32 ≈ 2.5 times lower than $\tilde{j}_p$. Essentially, formula (34) implies a rise in the "breakdown voltage" of the dynamic *p–n* junction with expanding diffusion layer and decreasing hole density in the space charge region. This effect is well documented for ordinary *p–n* junctions [15, 16, 18], where the positive space charge is due to donors rather than to holes.

At $\tilde{j}_p > \tilde{j}_b$, avalanche multiplication of holes starts in the anodic space charge region; however, the associated decrease in velocity v_p^f cannot be more than $\mu_e(\bar{v}_e + \bar{v}_h)/\bar{v}_e(\mu_e + \mu_h) \approx 1.7$ times. Otherwise, the condition $(p - N) > n$, which necessarily follows from the Poisson equation, is violated in the part of the space charge region that is adjacent to the anode front. Nevertheless, the rate of growth of w_p and, hence, of the voltage drop across the space charge region decreases considerably. In addition, so does the average space charge density, since avalanche multiplication causes the redistribution of the electron and hole concentrations in the space charge region [18], which also favors a drop of V_m. Thus, impact ionization appreciably slows down the increase in the voltage drop across the anodic space charge region, while to a considerably smaller extent than across the cathodic space charge. Numerical simulation shows that, when the current interrupted by the device described at the end of Sect. 3 rises, for instance, twofold, the amplitude of the voltage pulse increases by a factor of 1.5 (Fig. 4). This means that such an important quality index of the interrupter as the overvoltage coefficient [5] must be maximal at $\tilde{j}_p \approx \tilde{j}_b$ and $V_m \approx V_b$, as readily follows from Fig. 4a. Certainly, a decrease in the plasma boundary velocity due to impact ionization extends the pulse rise time (see Fig. 4b).

Another important parameter is peak power $P_m = V_\mathrm{m}^2/R$ applied to the load, which reaches maximal value P_0 at certain load resistance $R = R_0$ [5]. Basically, R_0 can be estimated by formula (30). However, a more accurate estimate is obtained from position ζ_0 of an extremum of the ζ dependence of u_m^2/χ, this position being found by numerically solving system of equations (24)–(25) when ζ/χ = const. It turned out that, in the range $10^{-2} < \zeta/\chi < 10^2$, the desired quantity differs from a solution to the equation

$$\zeta_0(1 + 25\chi/\zeta_0)^{1/4} = 1 \tag{35}$$

by no more than 10%. Moreover, the relationships

$$e_\mathrm{m} = \sqrt{\chi/25\zeta_0}, \quad u_\mathrm{m} = \frac{3}{2}e_\mathrm{m}^{3/2}, \tag{36}$$

are valid throughout extremal curve (35) within the same error. From (36), one can find resistance R_0 and interrupted current density $\tilde{j}_0$ at which the peak power is maximal and $E_p^\mathrm{m} = E_\mathrm{b}$. It should be taken into account that usually $\chi/\zeta_0 > 1$ (see below); therefore, solution (35) can be represented in the form $\zeta_0 = (25\chi)^{-1/3}$. Upon straightforward transformations, we get

$$R_0 = m\sqrt{2\xi E_\mathrm{b}L/q\tilde{p}_pSm}, \tag{37}$$

$$\tilde{j}_0 = (50q\varepsilon v_\mathrm{h}E_\mathrm{b}^2\lambda_p\tilde{p}_pm/\xi SL)^{1/3}. \tag{38}$$

Using (36), it is easy to check that the voltage amplitude corresponding to these values of the load resistance and interrupted current is equal to

$$V_0 \approx \frac{3}{2}mE_\mathrm{b}(\varepsilon v_\mathrm{h}\lambda_pS)^{1/3}\left(\frac{\xi E_\mathrm{b}L}{qp_pSM}\right)^{1/6}, \tag{39}$$

and the peak power, to

$$P_0 \approx 2mE_\mathrm{b}^2(\varepsilon v_\mathrm{h}\lambda_pS)^{2/3}\left(\frac{q\tilde{p}_pSm}{\xi E_\mathrm{b}L}\right)^{1/6}. \tag{40}$$

Rise time t_0 of the pulse forming under these conditions can be evaluated by using the approximation $\theta_{0.9} = 2(\chi/\zeta_0)^{1/4}$. In the interval ζ/χ = 0.1–10 of interest on extremal curve (35), this approximation provides a higher accuracy than (29). Substituting (37) and (38) into the extremal curve gives

$$t_0 \approx \sqrt{\frac{2q\tilde{p}_pSL}{5\xi E_\mathrm{b}m}} = \sqrt{\frac{4}{5}\frac{L}{R_0}}. \tag{41}$$

For the example considered at the end of Sect. 3, the calculation by these formula yields $R_0/m \approx 1.85\ \Omega$, $j_0 \approx 3.93$ kA/cm^2, $V_0/m \approx 523$ V, $P_0/m \approx 148$ kW, and $t_0 \approx$ 4.35 ns. Parameter χ is then equal to

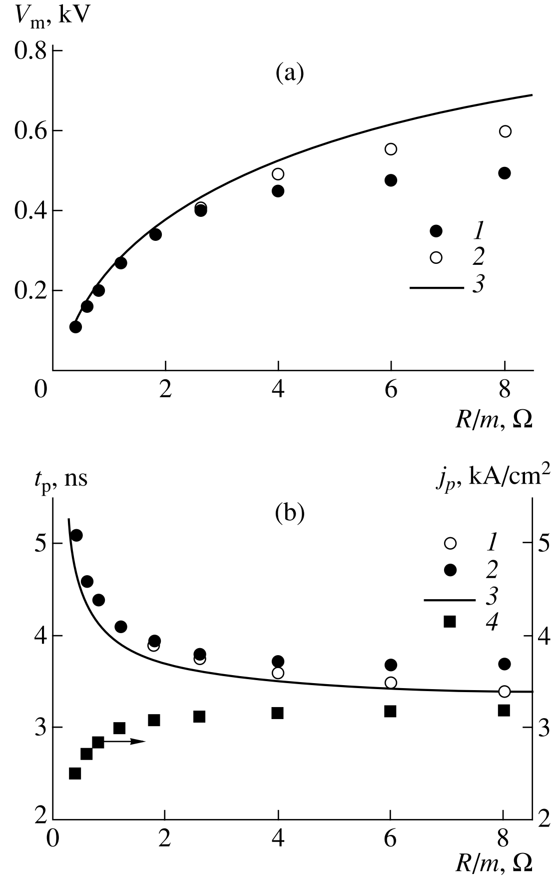


Fig. 5. Dependences of the (a) amplitude, as well as (b) rise time of the voltage pulse in the anodic space charge region and the interrupted current density on the load resistance at $\tilde{p}_p = 2 \times 10^{16}$ cm^{-3} with voltage $U_{\mathrm{c}0}$ fixed. Numerical simulation (*1*) with and (*2*) without regard to impact ionization in the anodic space charge region, (*3*) calculation by formulas (30) and (31) with regard to a change in the interrupted current density, and (*4*) interrupted current density obtained by numerical simulation.

$$\chi_0 = \frac{\varepsilon v_\mathrm{h}}{\lambda_p^2}\sqrt{\frac{\xi E_\mathrm{b}SL}{2q\tilde{p}_pm}} \approx 0.64, \tag{42}$$

and $\chi_0/\zeta_0 \approx 1.6$. In practice, this ratio cannot be decreased considerably, which validates using the approximate solution to equation (35).

5. DISCUSSION

The theory of space charge formation under high-density current interruption, which has been developed in the preceding sections, is qualitatively adequate to the SOS effect observed experimentally. However, dependences (34) and (39) of the maximal amplitude of voltage pulses on depth l_p of the p–n junction are too weak to quantitatively explain the multiple increase in

V_m [5] when l_p increases only by a factor of 1.5–2.0. Additional considerations which follow may explain this effect quantitatively.

First, we assumed above that the basic parameters of the theory are independent of each other and so obtained the relationships between them in the "pure" form. For example, formula (30) describes the dependence $V_m(l_p)$ at constant $\tilde{j}_p$ and $\tilde{p}_p$. However, all other parameters of the interrupter and circuit being the same, a change in l_p is bound to alter (i) the voltage drop across the diodes during time T^+ of the forward current pulse and, hence, the charge passing through the diodes; (ii) the total charge of nonequilibrium holes accumulated in the base layers of the diodes during time T^+ and hole distribution $p(x)$; (iii) the instant T_p the anodic space charge starts to form and corresponding current density $\tilde{j}_p$; and (iv) the position of the plane $x = \tilde{x}_p$, where the anodic space charge starts to form, and corresponding concentration $\tilde{p}_p$ of nonequilibrium holes.

In other words, parameters λ_p, $\tilde{j}_p$, and $\tilde{p}_p$ are interrelated in a rather complex manner. As a result, parameters $\tilde{j}_p$ and $\tilde{p}_p$ are very difficult to keep constant as λ_p varies in the course of the experiment (even numerical) especially if such a goal is not on hand. This is why, when interrupters of different types are tested under "identical conditions" (see, e.g., in [5, Fig. 24]), some of them can operate in the optimal regime, whereas others operate under conditions far from optimal. In view of the aforesaid, such a comparison seems to be not quite correct.

Second, as l_p grows (all other things being the same), the thickness of the weakly doped base and, hence, propagation losses of the forward current pulse decrease [7]. However, it is not necessary to increase l_p in order to reduce these losses: the same effect can be reached by shrinking the structure as a whole.

Third, as l_p grows, usually so does time interval T_p between the instant the current changes sign and the instant of current interruption.[4] This effect greatly favors the achievement of a desired value of the interrupted current when interrupters based on deep *p*–*n* junctions are used and is an important advantage of these interrupters over standard diodes.

Finally, it was assumed by default that all the diodes of the interrupter are identical and their parameters are uniformly distributed over the surface area. Actually, however, there always exists a spread of electrophysical parameters both within a device and from device to device. This inevitably causes a spread in the times of anodic space charge formation in different diodes, slows down the process of current interruption, and, consequently, decreases the amplitude of the voltage pulse across the load. The spread of the parameters can be expected to diminish with an increase in the diffusion depth, all other things being equal. This may be one more important process-related (rather than physical) factor responsible for the performance improvement of interrupters consisting of many series-connected diodes, which is observed with increasing l_p [5]. From the above, it follows that verification of the theoretical results needs additional experiments with as few carefully selected series-connected diodes as possible, as well as consideration and exhaustive control of all operating parameters of the interrupter.

In summary, we note that the diode structure shown in Fig. 1 was taken as an example, because it was studied earlier [7]. Also, it enables us to distinctly discriminate the stages of formation of the cathodic and anodic space charge regions. However, from the practical standpoint, such discrimination is rather a serious disadvantage, since the anodic space charge region, not participating in the formation of the basic pulse, produces a "pedestal" in front of the pulse and so results in useless energy losses for time period $T_n - T_p$. Obviously, this disadvantage can be eliminated by changing the doping profile so that the equality $T_p = T_n$ is satisfied most accurately. In essence, the requirement that the space charge regions begin to form simultaneously is equivalent to the condition necessary for the efficient operation of drift step-recovery diodes: the plasma fronts must simultaneously arrive at the plane of the *p*–*n* junction [20]. From Sect. 2, it follows that basically this can be achieved by changing the relationship between the depths of the n^+–n and *n*–*p* junctions.

ACKNOWLEDGMENTS

The authors are grateful to S.N. Yurkov and T.T. Mnatsakanov for assistance in simulation with the Issledovanie program.

This work was supported by the Russian Foundation for Basic Research.

REFERENCES

1. H. Benda and E. Spenke, Proc. IEEE **55**, 1331 (1967).
2. H. Benda and F. Dannhauser, Solid-State Electron. **11**, 1 (1968).
3. S. N. Rukin and S. P. Timoshenkov, in *Proceedings of the 9th Symposium on High-Current Electronics, Russia, 1992,* pp. 218–219.
4. S. A. Darznek, Yu. A. Kotov, G. A. Mesyats, and S. N. Rukin, Dokl. Akad. Nauk **334**, 304 (1994) [Phys. Dokl. **39**, 105 (1994)].
5. S. N. Rukin, Prib. Tekh. Éksp., No. 4, 5 (1999).
6. S. A. Darznek, G. A. Mesyats, and S. N. Rukin, Zh. Tekh. Fiz. **67** (10), 64 (1997) [Tech. Phys. **42**, 1170 (1997)].

---

[4] At relatively low current densities, such a situation was observed experimentally [19]; for high currents, it directly follows from formula (11).

7. S. A. Darznek, S. N. Rukin, and S. N. Tsyranov, Zh. Tekh. Fiz. **70** (4), 56 (2000) [Tech. Phys. **45**, 436 (2000)].
8. S. N. Rukin and S. N. Tsyranov, Pis'ma Zh. Tekh. Fiz. **26** (18), 41 (2000) [Tech. Phys. Lett. **26**, 824 (2000)].
9. A. V. Ponomarev, S. N. Rukin, and S. N. Tsyranov, Pis'ma Zh. Tekh. Fiz. **27** (20), 29 (2001) [Tech. Phys. Lett. **27**, 857 (2001)].
10. S. N. Rukin and S. N. Tsyranov, Pis'ma Zh. Tekh. Fiz. **30** (1), 43 (2004) [Tech. Phys. Lett. **30**, 19 (2004)].
11. A. V. Gorbatyuk and P. V. Rodin, Preprint No. 1276, FTI (Ioffe Physicotechnical Institute, Leningrad, 1988).
12. A. F. Kardo-Sysoev and M. V. Popova, Fiz. Tekh. Poluprovodn. (Leningrad) **25**, 3 (1991) [Sov. Phys. Semicond. **25**, 1 (1991)].
13. F. N. Trofimenkoff, Proc. IEEE **53**, 1765 (1965).
14. Yu. N. Serezhkin and A. A. Shesterkina, Fiz. Tekh. Poluprovodn. (St. Petersburg) **37**, 1109 (2003) [Semiconductors **37**, 1085 (2003)].
15. S. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981; Mir, Moscow, 1984).
16. A. S. Kyuregyan and S. N. Yurkov, Fiz. Tekh. Poluprovodn. (Leningrad) **23**, 1819 (1989) [Sov. Phys. Semicond. **23**, 1126 (1989)].
17. T. T. Mnatsakanov, I. L. Rostovtsev, and N. I. Philatov, Solid-State Electron. **30**, 579 (1987).
18. I. V. Grekhov and Yu. N. Serezhkin, *Avalanche Breakdown of p–n Junction in Semiconductors* (Énergiya, Leningrad, 1980) [in Russian].
19. I. V. Grekhov, E. M. Geifman, and L. S. Kostina, Zh. Tekh. Fiz. **53**, 726 (1983) [Sov. Phys. Tech. Phys. **28**, 459 (1983)].
20. I. V. Grekhov, Izv. Ross. Akad. Nauk, Énerg., No. 1, 53 (2000).

Translated by M. Lebedev

Technical Physics, Vol. 50, No. 7, 2005, pp. 914–917. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 97–100.
Original Russian Text Copyright © 2005 by Svalov, Vas'kovskiy, Schegoleva, Kurlyandskaya.

SOLID-STATE ELECTRONICS

Effect of the Layer Thickness on the Magnetic Properties and Structure of Terbium in $(Tb/Ti)_n$ and $(Tb/Si)_n$ Multilayer Films

A. V. Svalov[1], V. O. Vas'kovskiy[1], N. N. Schegoleva[1], and G. V. Kurlyandskaya[1, 2]
[1] *Gorky Ural State University, pr. Lenina 51, Yekaterinburg, 620083 Russia*
e-mail: andrey.svalov@usu.ru
[2] *Universidad de Oviedo, Oviedo, 33007 Spain*

Received August 23, 2004; in final form, November 10, 2004

Abstract—The bulk and surface structures and the magnetic properties of Tb layers in $(Tb/Ti)_n$ and $(Tb/Si)_n$ multilayer films are studied experimentally. As the magnetic layer thickness decreases, Tb becomes amorphous. Along with the amorphization, the magnetic ordering temperature declines and the temperature range of magnetic hysteresis shifts. In terms of the ZFC–FC approach, this shift means crystalline magnetic anisotropy breaking. The material of nonmagnetic spacings between the Tb layers plays a certain role in these changes. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Low-dimensional objects occupy a prominent place in contemporary research on magnetism. Such objects are, in particular, thin magnetic films, among which those of rare-earth metals stand out. These films provide information on the fundamental properties of thin-film 4*f* magnets, including information on indirect exchange interaction in them, [1, 2] and are promising materials for multilayer structures [3, 4]. Gadolinium films (gadolinium is known to have the highest Curie temperature among rare earths) have attracted the greatest attention [5–7]. The magnetic ordering temperature of terbium is slightly lower; however, unlike gadolinium, terbium atoms have a nonspherical electron shell. This feature is a prerequisite for high magnetic anisotropy, which makes Tb-containing film very promising for applications. In this work, we study a relation between the structure and magnetic properties of Tb layers incorporated into $(Tb/Ti)_n$ and $(Tb/Si)_n$ multilayer films.

EXPERIMENTAL

The films were deposited by rf ion sputtering on glass substrates at a deposition rate of ≈0.1 nm/s. The working chamber was preevacuated to a pressure of 1×10^{-6} Torr, and the argon pressure in the discharge was 2×10^{-4} Torr. The test samples were prepared by alternatively depositing Tb layers of different thicknesses (h_{Tb} = 2.5–600 nm) and 2-nm-thick nonmagnetic (Ti or Si) spacers. The total thickness of the magnetic component in the multilayer samples was constant (≈200 nm). Each sample had a buffer sublayer and an insulating nonmagnetic coating.

The magnetic properties of the films were measured with a torquemeter in the temperature range 80–300 K. The thickness and surface morphology of the films were examined in an atomic force microscope (AFM). The crystal structure was examined under an electron microscope on thin (50 nm) samples deposited onto the cleavage surface of NaCl crystals.

RESULTS AND DISCUSSION

Figure 1 shows the dependences of torque L on temperature T for the (*a*) single-layer Tb film with h_{Tb} = 600 nm, as well as for (*b*–*d*) $(Tb/Ti)_n$ and (*e*, *f*) $(Tb/Si)_n$ multilayer films with different thicknesses of the magnetic layers. The $L(T)$ curves are constructed for two initial states of the samples. These states set in when the sample is cooled in the absence (zero-field cooling (ZFC)) and in the presence (field cooling (FC)) of a 1-kOe magnetic field making an angle of 45° with the film plane (Fig. 1; curves *2* and *1*, respectively). Torque L was measured in the same magnetic field.

As is seen from Fig. 1, most of the samples exhibit a strong thermomagnetic hysteresis of torque, which is observed at $T < 200$ K and reflects, in fact, the temperature hysteresis of magnetization. This feature (which is also typical of spin glasses [8]) is likely to be associated with a high crystalline magnetic anisotropy of terbium at low temperatures [9]. Upon cooling in the magnetic field, the disordered magnetic state of the Tb layers (the FC mode) turns into the magnetic texture, which specifies a remanent magnetization and a torque. The ZFC mode causes the demagnetized state. Below temperature T_f at which the difference between curves *1* and *2* becomes appreciable, the magnetic field used in the experiment is too low to effectively magnetize the samples. Therefore, their magnetization and, correspondingly, torque are low.

Remarkably, temperature T_f depends on the magnetic layer thickness and the material of the nonmagnetic spacers. The dependence $T_f(h_{\text{Tb}})$ (Fig. 2, curve *1*) demonstrates that the thermomagnetic hysteresis shifts toward lower temperatures as the magnetic layer thickness decreases with the introduction of nonmagnetic Ti spacers. For the $(\text{Tb}(2.5)/\text{Ti})_{60}$ sample, this hysteresis is absent throughout the temperature range under study (Fig. 1, curve *d*). At the given Tb layer thickness (h_{Tb} = 2.5 nm), T_f is likely to drop below 80 K. The $(\text{Tb/Si})_n$ films behave in a similar way: the only difference is that the temperature hysteresis weakens even at h_{Tb} = 10 nm (Fig. 1, curve *f*).

Figure 2 (curves *2*, *3*) also shows the dependences of magnetic ordering temperature T_{ord} on the Tb-layer thickness. Temperature T_{ord} was estimated from the $L(T)$ curves by linear extrapolation of the high-temperature ($T > T_f$) portions to the zero value of L. Although the error inherent in this procedure is high (±5), the data presented suggest that T_{ord} is constant over a wide h_{Tb} range. However, when the Tb layer thickness is less than 10 nm in the $(\text{Tb/Ti})_n$ samples or less than 20 nm in the $(\text{Tb/Si})_n$ samples, T_{ord} decreases substantially. A similar dependence found earlier in layered Gd films when the magnetic layers were thinned [7] was attributed to a buildup of lattice microstrains, which eventually cause amorphization of the films. The same correlation between the structure and thickness of the Tb layers was established in layered Tb films.

Figure 3 shows the electron diffraction patterns and microstructures of two $(\text{Tb/Si})_n$ films with a Tb layer thickness of (Fig. 3a) 5 and (Figs. 3b, 3c) 20 nm. Their structural characteristics are seen to be different. The electron diffraction pattern taken from the sample with thin Tb layers does not contain distinct lines, exhibiting a ripple in the dark-field image (Fig. 3a). This is a clear indication of the amorphous state. The second sample is in the polycrystalline state, as follows from a set of clear-cut diffraction lines and bright spots (the latter are due to coherent domains) in the dark-field image. Analysis showed that such a diffraction pattern can be related to the hcp crystal lattice of Tb. The crystallite size is typically 5–10 nm.

Thus, for both the Tb and Gd systems, the fine-grained state turns into the amorphous state with decreasing layer thickness. The radical structural transformations occur for h_{Tb} ranging from 20 to 5 nm. As was shown above, the magnetic properties also change significantly at these thicknesses (specifically, T_{ord} decreases). It can thus be concluded that amorphization effectively suppresses exchange interaction. This finding is not surprising: atomic spacing fluctuations are superimposed on the exchange interaction energy of rare-earth metal, which oscillates with distance, and, thereby, frustrate the system of atomic magnetic moments. However, such behavior distinguishes Tb films from Gd ones, in which T_{ord} decreases noticeably even in the crystalline phase at a sufficiently high level of defects [7].

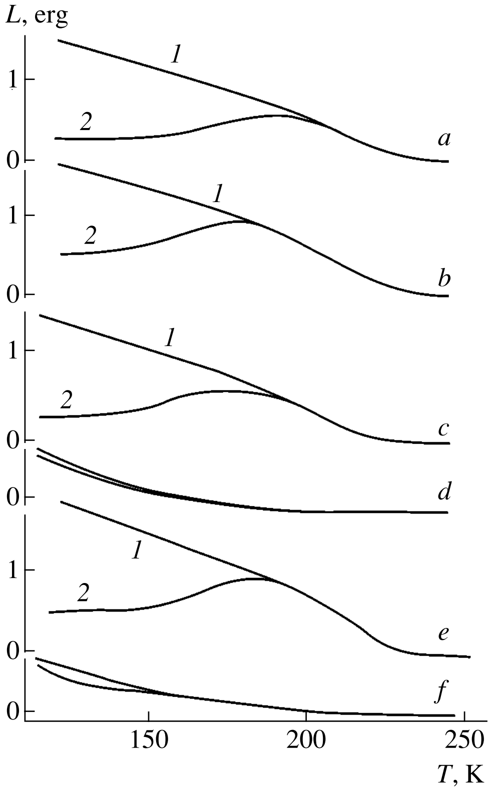


Fig. 1. Temperature dependences of the torque for (a) the single-layer Tb film 600 nm thick, as well as for the (b–d) $(\text{Tb/Ti})_n$ and (e, f) $(\text{Tb/Si})_n$ multilayer films with magnetic layers thickness h_{Tb} = (b, f) 10, (c) 5, (d) 2.5, and (e) 20 nm. The curves are obtained under the (*1*) FC and (*2*) ZFC conditions.

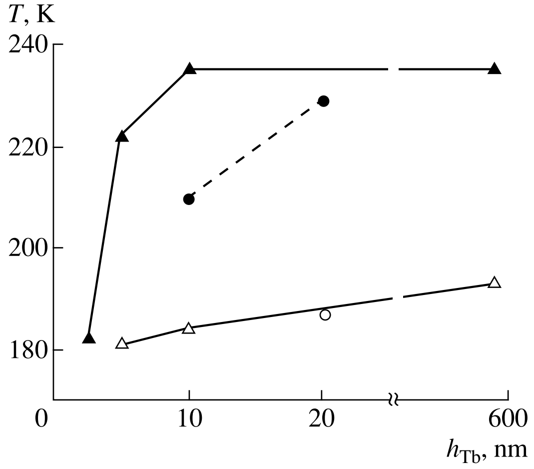


Fig. 2. Dependences of (▲, ●) magnetic ordering temperature T_{ord} and (△, ○) temperature T_f of onset of thermomagnetic hysteresis on the Tb layer thickness in the (▲, △) $(\text{Tb/Ti})_n$ and (●, ○) $(\text{Tb/Si})_n$ multilayer films.

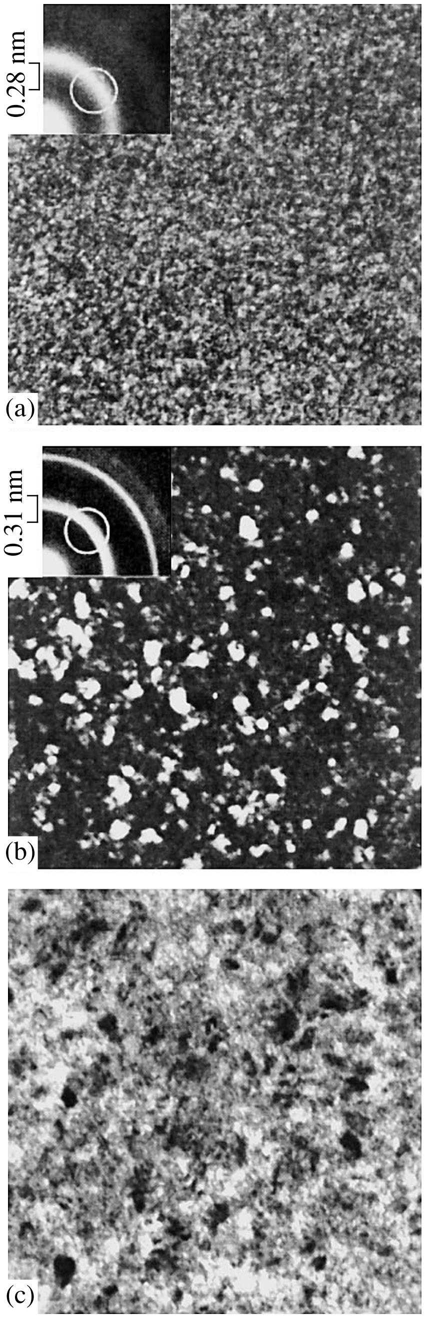


Fig. 3. (a, b) Dark- and (c) bright-field electron microscopic images of the $(Tb/Si)_n$ film microstructure and electron diffraction patterns for h_{Tb} = (a) 5 and (b, c) 20 nm. The circles show the positions of the diaphragm in dark-field imaging. The figures are the interplanar spacings corresponding to the centers of the diffraction reflections selected.

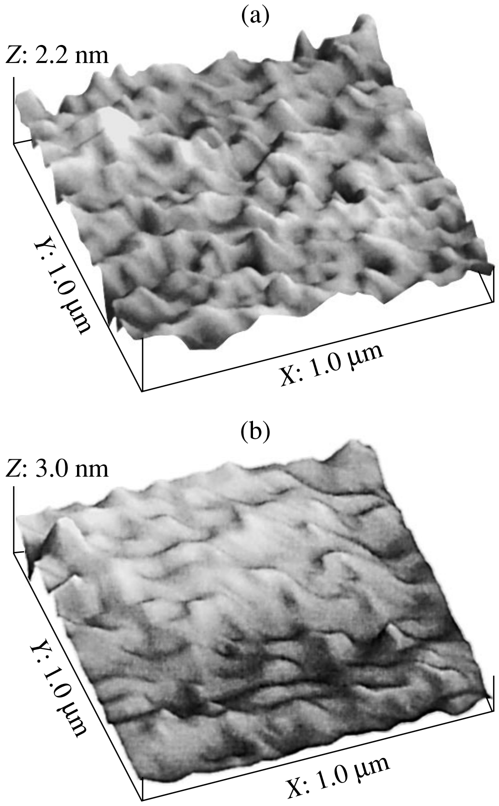


Fig. 4. AFM surface images of the (a) $(Ti/Tb(10\ nm))_{20}$ and (b) $(Si/Tb(10\ nm))_{20}$ samples and surface roughness dispersion R_{rms}.

Temperature T_f, as well as T_{ord}, also seems to be a structure-sensitive parameter. Its decrease upon amorphization points to a correlation between the thermomagnetic hysteresis and crystalline magnetic anisotropy, which breaks under the structural transformation. There is evidence that a number of terbium-based amorphous alloys (Tb–Cr [10], $Tb_{90}Si_{10}$ [11], and Tb [12]) behave as spin glasses. However, characteristic T_f temperature for them is low, ≈50 K. In the works cited above, the thermomagnetic hysteresis is related to the dispersion of the local magnetic anisotropy of Tb ions, which arises because of their nonspherical 4*f* electron subshell. The temperature dependences of the local magnetic anisotropy parameters are likely to differ from the temperature dependence of the macroscopic magnetic anisotropy, which is responsible for thermomagnetic hysteresis in the polycrystalline Tb layers. As a result, temperature T_f typical of the amorphous state is substantially different (is much lower).

As was noted above (see also Figs. 1 and 2), the material of the nonmagnetic spacers is to some extent responsible for the thickness variation of the magnetic properties of the Tb layers. Specifically, in the films with Si spacers, a trend toward magnetism degradation and crystalline magnetic anisotropy breaking shows up at h_{Tb} that is slightly higher than in the Ti-containing films. This may be due to interlayer diffusion and the formation of transition regions with a variable composition. Silicon, being an effective amorphizer, favors the degradation of the Tb crystalline state to a greater extent than Ti. Next, the depths of penetration of different nonmagnetic atoms into Tb or the morphologies of different nonmagnetic spacers may also differ. Relevant information can be obtained by analyzing the surface structure of the films on the assumption that it is related to the state of the interfaces.

Figure 4 shows the AFM images of the surfaces of the $(Tb(10\ nm)/Ti)_{20}$ and $Tb(10\ nm)/Si)_{20}$ films. These surfaces are seen to differ considerably: the surface of the sample with the Si layers contains larger features compared with the sample with the Ti layers. Dispersion R_{rms} of the surface roughness (Fig. 4) for these samples is also different: the surface of the $(Tb(10\ nm)/Ti)_{20}$ sample is seen to be smoother. These differences may be viewed as an indication of effective Si–Tb mixing at the interfaces, which results in more effective amorphization of the rare earth metal.

CONCLUSIONS

Thus, the fine-grained state of Tb entering into $(Tb/Ti)_n$ and $(Tb/Si)_n$ multilayer films turns into the amorphous state as the magnetic layer thickness decreases below $h_{Tb} < 20$ nm. This transition decreases the magnetic ordering temperature and also suppresses thermomagnetic hysteresis; this, in turn, points to crystalline magnetic anisotropy breaking. The amount of the structural and magnetic transformations depends on the material of nonmagnetic spacers between Tb layers. Silicon favors amorphization and changes the magnetic properties of Tb as the magnetic layers get thicker.

ACKNOWLEDGMENTS

We thank I. Goikuria Orue (Universidad del Pais Vasco, Bilbao, Spain) for the assistance in atomic force microscopy studies.

This work was supported by the Russian Foundation for Basic Research (project no. 04-02-16485a), the program “Russian Universities” (project no. ur. 01.01.060), and the program “Ramon y Cajal” of the Ministry of Science and Technology of Spain.

REFERENCES

1. C. Dufour, K. Dumesnil, A. Mougin, *et al.*, J. Phys.: Condens. Matter **9**, L131 (1997).
2. A. T. Hindmarch and B. J. Hickey, Phys. Rev. Lett. **91**, 116601 (2003).
3. S. Tsunshima, J. Magn. Magn. Mater. **156**, 283 (1996).
4. Z. S. Shan and D. J. Sellmyer, in *Handbook on the Physics and Chemistry of Rare Earths*, Ed. by K. A. Gschneidner, Jr. and LeRoy Eyring (North-Holland, Amsterdam, 1996), Vol. 22, p. 81.
5. C. F. Majkrzak, J. W. Cable, J. Kwo, *et al.*, J. Appl. Phys. **61**, 4055 (1987).
6. M. Farle, Rep. Prog. Phys. **61**, 755 (1998).
7. V. O. Vas’kovskiy, A. V. Svalov, A. V. Gorbunov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **43**, 672 (2001) [Phys. Solid State **43**, 698 (2001)].
8. C. Y. Huang, J. Magn. Magn. Mater. **51**, 1 (1985).
9. S. A. Nikitin, *Magnetic Properties of Rare-Earth Metals and Alloys* (Mosk. Gos. Univ., Moscow, 1989) [in Russian].
10. O. V. Stognei and Ö. Rapp, J. Magn. Magn. Mater. **196–197**, 266 (1999).
11. J. J. Hauser, Phys. Rev. B **34**, 3212 (1986).
12. J. J. Hauser, Solid State Commun. **55**, 163 (1985).

Translated by K. Shakhlevich

Technical Physics, Vol. 50, No. 7, 2005, pp. 918–923. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 101–105.
Original Russian Text Copyright © 2005 by Sankin, Malykh.

ACOUSTICS, ACOUSTOELECTRONICS

Force Acting on a Cylinder under Ultrasonically Induced Cavitation

G. N. Sankin* and N. V. Malykh**

** Lavrent'ev Institute of Hydrodynamics, Siberian Division, Russian Academy of Sciences,*
pr. Akademika Lavrent'eva 15, Novosibirsk, 630090 Russia

e-mail: sankin@hydro.nsc.ru

*** Kutateladze Institute of Thermophysics, Siberian Division, Russian Academy of Sciences,*
pr. Akademika Lavrent'eva 1, Novosibirsk, 630090 Russia

e-mail: malykh@itp.nsc.ru

Received July 20, 2004

Abstract—To elucidate mechanisms of cavitation action on a surface with microcapillary discontinuity and refine the model of the sonocapillary effect, the force acting on a cylinder is correlated with the height of a liquid in the capillary under cavitation, both quantities being measured at the same point of the ultrasonic field. It is found that the force acting on the cylinder is directed toward a cavitation cluster stabilized at the end face of the cylinder. This force can be enhanced with another cylinder placed on the side of the cluster coaxially with the first one. The dynamics of the cavitation cluster depending on the cylinder spacing is investigated in the case when one of the cylinders is on a pendulum suspension. The conditions for self-sustained oscillations of the pendulum are found. The dependence of the attracting force between the cylinders on the ultrasonic frequency and cylinder spacing is derived. The end-face-averaged pressure exerted on a cylinder of diameter 1.2 mm may reach 0.16 kPa, and the intracapillary pressure attains 0.89 kPa. Thus, the sonocapillary effect may be explained, at least partially, by a counterpressure arising from cluster–capillary interaction. This effect can be used in designing cavitation sensors and ultrasonic sources. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

A body placed in an acoustic field experiences the action of ponderomotive forces, those that are proportional to the sound amplitude and those that depend quadratically on the radiation pressure, and also Stokes and Bernoulli hydrodynamic forces. However, when the ultrasonic intensity is high (more than 1 W/cm^2), the basic action on a solid surface is due to cavitation effects, such as shock waves generated by spherically symmetric and cumulative collapse of bubbles or their clusters, the wedging effect of bubbles, and the sonocapillary (SC) effect in voids of solids [1]. The SC effect, a rise of a liquid in a capillary under ultrasonic action, has long attracted the attention of researchers [2] and offers a variety of applications, from purification to dispersion of a wide class of materials [3].

In a gap between vibrating walls, hydrodynamic vortices [4] and cavitation bubble clusters [1, 5] may form. When cumulatively collapsing (spherically or with the formation of a stream), bubbles and their clusters generate short (compared with the ultrasound period) pressure pulses, which cause erosion [5, 6]. Similar phenomena (vorticity and cavitation) that give rise to the SC effect and exert a pressure on a solid wall were observed at the end face of a solid rod and capillary placed in an ultrasonic field [7]. Measurement of a pressure exerted on the end face of a capillary under cavitation might be a decisive argument for one or another model of the SC effect. The point of application of the counteracting force influencing the liquid flow in the capillary is yet to be found.

In this work, we study the behavior of the cluster-cavitation-induced force acting on a solid cylinder placed in an ultrasonic field.

EXPERIMENTAL

The experimental setup is shown in Fig. 1a. A standing ultrasonic wave is generated by spherical piezoelectric transducer *1* placed in a vessel with a liquid under study. The transducer, made of silver-plated PZT-19 ceramics of diameter 50 mm, is connected to specially designed rectangular pulse generator, which generates pulses of amplitude 90–120 V and repetition rate f = 26–31 kHz, through a variable inductance in order to tune to resonance and generate a sine-shaped signal. The signal waveform was monitored with a Tektronix TDS-210 oscilloscope connected to the source (radiator) through a divider.

The sphere with three observation slits *10* is mounted on rubber support *9*, which is placed into 1-l glass beaker *2* filled with water. The water is first heated to reduce the gas concentration in it and then is cooled to 35–45°C. The water thus prepared can be used in experiments for several hours. Level *8* of the liquid is above the top of the source by $h = 19–29$ mm. Since the standing wave generated in the beaker has an integer number of half-waves fitting the liquid column, resonance frequency f_r depends on h.

Hinged (*5*) pendulum *3* with a cylinder made of copper wire at its end is freely suspended above the source. The wire is insulated by a lacquer. Cylinders I and II of diameters $d = 1.2$ and 2.2 mm, respectively, are shown in Fig. 1b.

A cavitation cluster near the end faces of cylinders *3*' and *7*' was stabilized by the technique described in [7]. Due to cavitation, the pendulum was displaced from the equilibrium position by angle $\Delta\alpha$, which was determined from the displacement of the cylinder. Irrespective of whether both the end and lateral surfaces are filed or the end surface is filed while the lateral surface is covered by the lacquer, the cloud can be stabilized near the end face. This is because the cluster is more stable at the flat end face rather than at the convex wall of the cylinder.

If a cavitation cluster was not formed spontaneously at cylinder *3*', its formation was forced using wire *7*, which was alternately brought near the source and cylinder. The wire, which stabilizes (localizes) a bubble cluster [7], was put into the highest pressure area (the standing wave antinode) and then moved toward a desired place (specifically, toward the end face of the cylinder). When moving, the wire entrained the cluster.

It is known that the SC effect can be enhanced by pressing a capillary against the source [2]. To model this situation, we used auxiliary cylinder *7*' of diameter 1.05 mm, which was made of copper wire with the lacquer insulation stripped off. This cylinder could be placed symmetrically about the source axis and coaxially with cylinder *3*'.

The cylinder spacing was monitored visually, and the process was recorded by shadow photography using a SensiCam Fast Shutter camera (PCO, Kelheim, Germany) with a frame frequency of 30–150 per second (the exposure time is 1 ms). Illumination was accomplished with an incandescent lamp through clouded glass. To avoid distortions that can be introduced into the beaker image by the cylindrical wall, a 32 × 37 × 3-mm glass platelet was hermetically glued on the beaker and the gap between the wall and objective lens was filled with the same liquid as in the beaker to match the refractive indices.

From recorded deflection $\Delta\alpha$ of the pendulum in the terrestrial gravitational field, one can find force F acting on the cylinder. The system was calibrated by loads of known weights. The load-and-pulley arrangement was attached to the end of the pendulum suspended in air. The calibration curve is shown in Fig. 2. The curves are almost linear with slope K depending on the weight of the load and balance weight. It was found that $K =$ 28 µm/mg for $d = 1.2$ mm and $K = 35$ µm/mg for $d =$ 2.2 mm.

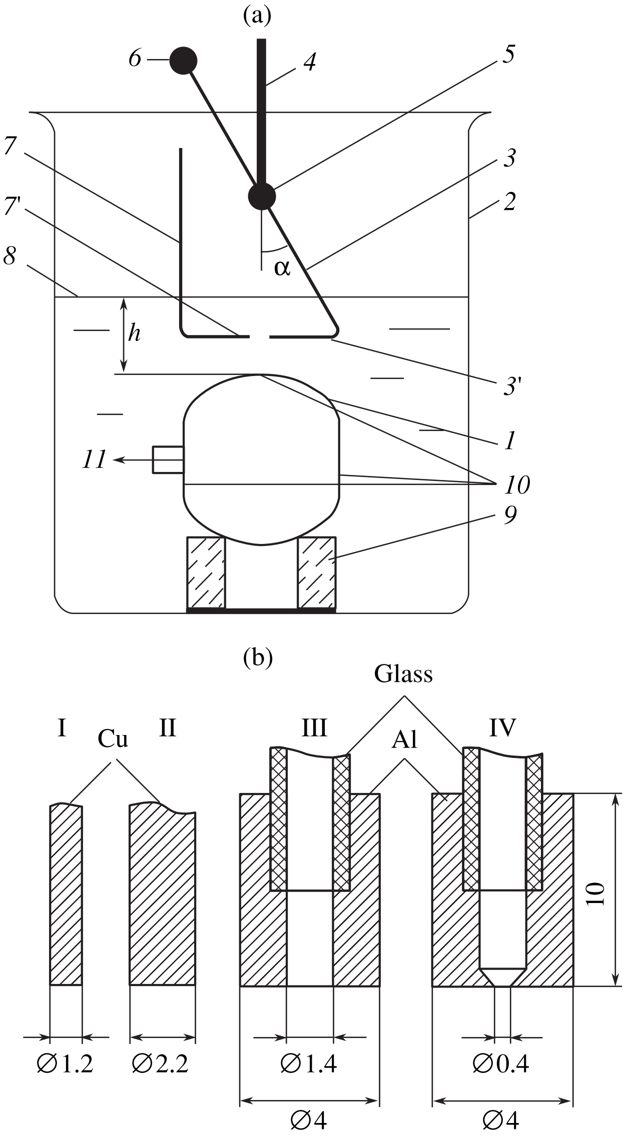


Fig. 1. (a) Design of the setup: *1*, piezoelectric ceramic sphere; *2*, beaker; *3*, pendulum; *7*, auxiliary wire; *3*' and *7*', cylinders; *4*, pendulum suspension; *5*, hinge; *6*, balance weight; *8*, free surface of liquid; *9*, rubber ring; *10*, observation slits; and *11*, ultrasonic generator. (b) Wires (I, II) and capillaries (III, IV) used in the experiments.

To correlate the force acting on the cylinder with the amount of the SC effect, we made an aluminum capillary with a transparent glass extension making it possible to observe the rise of the liquid (III and IV in Fig. 1b).

RESULTS

Visual observation shows that the cylinder is displaced only in the presence of a cavitation cluster near its end face.

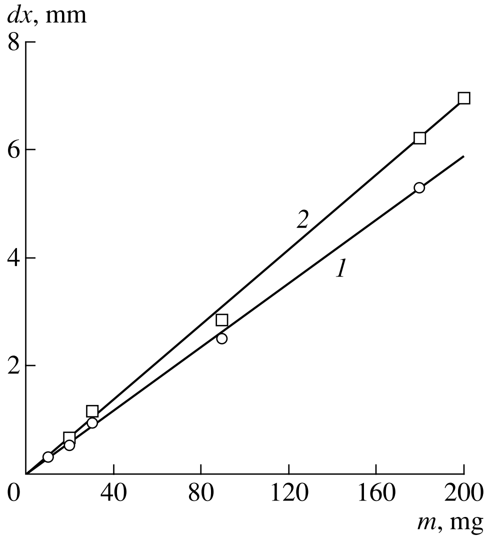


Fig. 2. Calibration curves for the pendulum with the cylinder of diameter (*1*) 1.2 and (*2*) 2.2 mm.

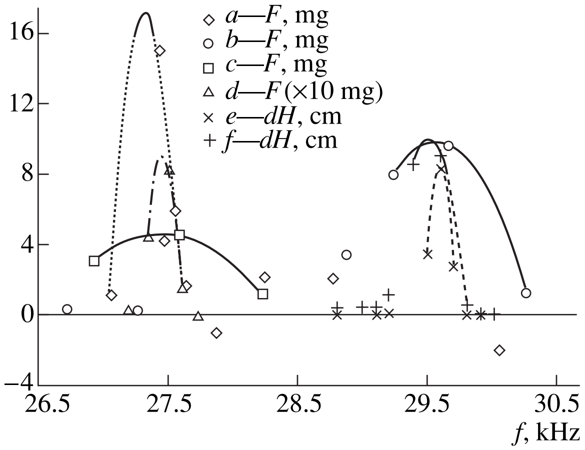


Fig. 3. (*a*–*d*) Force vs. ultrasonic frequency for (*a*–*c*) one and (*d*) two cylinders and (*e*, *f*) height of the liquid in the aluminum capillary vs. frequency. The cylinder diameter is (*a*) 1.2 and (*b*, *c*) 2.2 mm. The inner diameter of the capillary is (*e*) 0.4 and (*f*) 1.4 mm. The distance between the source and liquid surface h = (*a*, *c*, *d*) 19 and (*b*, *e*, *f*) 29 mm.

Digital filming made it possible to measure the force deflecting the pendulum from the equilibrium position. Figure 3 plots force F acting on the cylinder against ultrasonic frequency f. Each data point is the result of averaging over three measurements. The curve is of resonant character, the larger the diameter of the cylinder, the wider the resonance Fig. 3 (*b*, *c*). For the cylinder with a diameter of 2.2 mm, the resonance peak (h = 19 mm) is near f_r = 29.67 kHz; for h = 29 mm, at f_r = 27.58 kHz. For the smallest diameter cylinder, the force is the highest, about 15 mg (Fig. 3, curve *a*). Negative values of the force are due to the cluster stabilized at the lateral surface of the cylinder.

If second cylinder *7*' is placed x_0 distant from first cylinder *3*' (Fig. 4a), the cylinders are attracted to each other (Fig. 4b), thereby increasing the measured force acting on the right-hand cylinder. Figure 4 distinctly shows the formation of cavitation clusters near the end faces of the cylinders being attracted. At closer examination (Fig. 4b, the sixth frame from top), it was found that the force decreases when the cylinders come into contact with each other. The displacement of the pendulum was measured after its oscillations had decayed completely. The dependence of the attracting force between the cylinders on the ultrasonic frequency is shown in Fig. 3 (curve *d*). The curve is of resonant character with a peak at f_r = 27.47 kHz. At the maximum, the force reaches a value of 86 mg. The distance between the cylinders is calculated by the formula $x = x_0 - F*K$.

At a certain initial value of x_0 (with the ultrasonic source switched off, Fig. 4a), switching the source on causes self-oscillation of the cylinders lasting several minutes.

Figure 4b shows the variation of the gap and the shape of cavitation clusters near the end faces of both cylinders within an oscillation period. The cavitation-induced period-averaged force is seen to attract the cylinders to each other (the time-averaged position of the right-hand cylinder is shifted relative to its equilibrium position shown in Fig. 4a, $\langle x \rangle < x_0$. The left-hand pendulum remains almost unmoved, while the right-hand pendulum oscillates (Fig. 5a). The smaller the cylinder spacing, the greater the cavitation-induced attracting force between the cylinders. By the time the cylinders come into contact, the cloud is displaced from the gap, the attracting force decreases, and the right-hand pendulum tends toward the equilibrium position. However, it goes through the equilibrium position by inertia, strikes the left-hand pendulum, and bounces back from it. The gap widens, allowing for formation of a new bubble cluster, generating the attracting force, and the process recurs. Since the system possesses the feedback described above, self-sustained oscillations set in.

Figure 5b plots the rate of variation of cylinder spacing dx/dt against x (phase diagram) for more than two self-oscillation periods (ultrasonic frequency f = 27.5 kHz). The self-oscillation period is about 0.6 s, which corresponds to self-oscillation frequency f_{so} = 1.7 Hz ($f_{so} \ll f$).

The SC effect was studied with the aluminum capillary placed vertically, so that cavitation was generated at its end face. When the capillary and cylinder were immersed in the same area over the source, the level of the liquid in the extension under ultrasonically induced cavitation rose by dH (Fig. 3 (*e*, *f*)) compared with the no-cavitation case (the rise is due to capillary forces alone). It follows from Fig. 3 that the frequency dependences of the force acting on the cylinder and of the height of the liquid correlate: both peak near 29.6 kHz.

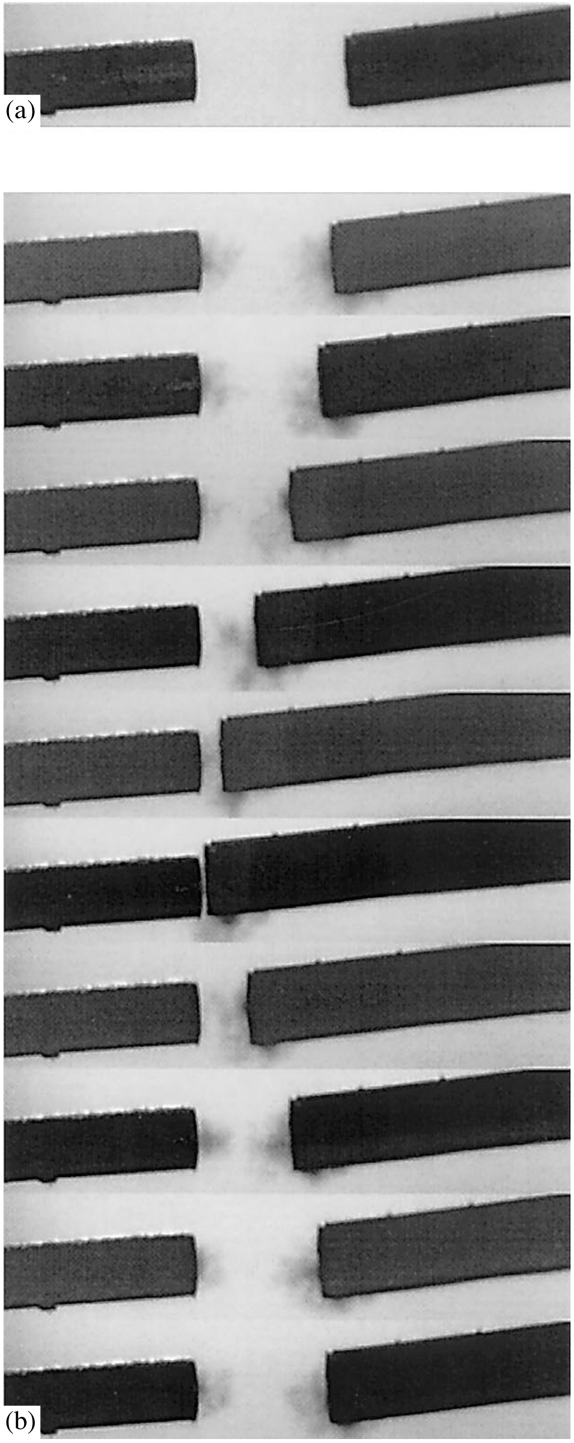


Fig. 4. Motion-picture frames of the 1.2-mm-diameter cylinder (on the right) attached to the pendulum and the auxiliary wire (on the left) when the ultrasonic source is switched (a) off and (b) on. Panel “b” shows cavitation clusters stabilized near the end faces. The exposure time is 1 ms; the time interval between frames, 67 ms.

The height of the liquid at resonance is almost independent of the capillary diameter and equals 8.4 and 9.1 cm for the smaller and larger diameters, respectively. The

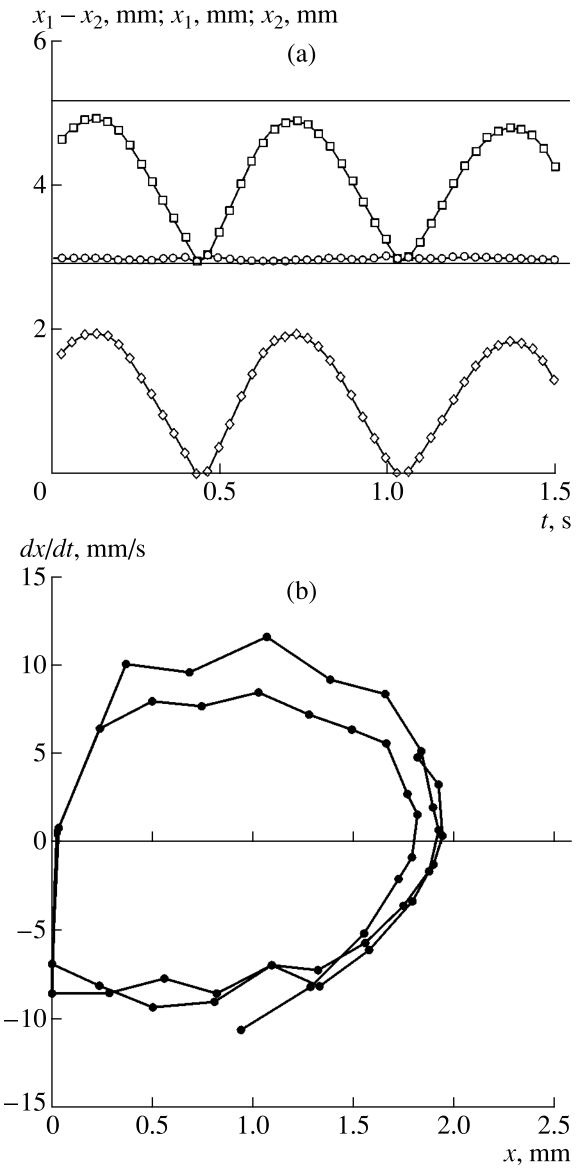


Fig. 5. (a) Self-oscillations of the (◇) gap, as well as of the coordinates of the (○) left- and (□) right-hand cylinders. The continuous curves refer to the equilibrium values of the coordinates of the end face (without ultrasound). (b) Phase diagram of the pendulum self-oscillations under cavitation. x_1 and x_2 are the coordinates of the left- and right-hand cylinders.

hydrostatic pressure of such a liquid column is about 0.89 kPa, which is 5.7 times higher than the pressure averaged over the cross-sectional area of the continuous cylinder of diameter $d = 1.2$ mm.

DISCUSSION

There are a number of models considering the SC effect. One of them relates the SC effect to the stream-generating collapse of the bubble near the mouth of the capillary [2]. If the bubble is on the capillary axis, one may expect that it will collapse nonspherically (nonuniformly), forming a cumulative stream. The stream entering the capillary imparts a momentum to the liquid, causing it to rise. The force acting on the capillary is directed oppositely to the flow velocity vector in the channel. Note that, generally, the formation of the stream is not obligatory. A shift of the center of mass of the bubbles due to the so-called Kelvin impulse [8] suffices to generate the SC effect. In this case, an increase in the spherical accumulation of the energy inside a bubble is consistent with experimental data for enhancement of the glow intensity from a cavitation cluster [9].

At a maximal rise of the liquid in the capillary, a cluster (and not a single bubble) was observed [7] at its end face that collapsed spherically symmetrically. Under cavitation, the net pressure field in the liquid is the sum of the field of the standing wave generated by the source inside the sphere and outside it (in the beaker) and the field of bubble-induced waves. In the rarefaction phase, the pressure does not drop below a value corresponding to the cavitation strength of the liquid. Transformation of the rarefaction wave into a compression wave (an analogue of the “ebullition” wave) and the radiation pressure also positively contribute to the pressure averaged over the wave period [10]. Thus, the pressure in the liquid under cavitation may oscillate about the mean exceeding the hydrostatic pressure. In this case, the force acting on the capillary is codirected with the flow velocity vector in the channel.

It was found that the force acting on the cylinder is directed toward a cavitation cluster stabilized at the end face of the capillary. It can be assumed that, when a cluster or a bubble collapses with the formation of a stream, the force acting on the liquid is counterbalanced by the force acting on the cylinder due to bubble–wall attraction. While not being completely adequate to the phenomenon considered, this model correctly predicts the force direction.

Since the pressure on the end face of the cylinder with a diameter of 1.2 mm reached 0.16 kPa at resonance, the SC effect can be explained, at least partially, by a counterpressure arising as a result of cluster–capillary interaction. The fact that the force increases in the presence of an opposite solid wall (Fig. 3d) strengthens this supposition. The maximal force with the second cylinder exceeds the force without it by a factor of 5.7. Note that the SC effect can be enhanced by tightly pressing the capillary against the wall of the source or cavity [2].

Therefore, we believe that the SC effect generalizes a number of nonlinear cavitation effects producing a positive mean pressure in the capillary. One can assume that the force acting on the end face of the cylinder placed at the center of the source will be much higher, since the SC effect is the most pronounced if the end face of the capillary is at the center of the sphere.

Modulation of the cavitation force acting on a cylinder by means of another cylinder generates self-sustained oscillations in the system. Self-oscillations set in when the initial oscillation amplitude exceeds the gap between the cylinders, thus providing the contact (collision) between them (Fig. 4). If collision is absent, the oscillations of the pendulum decay with a damping decrement depending, e.g., on the viscosity of the liquid.

CONCLUSIONS

Thus, it is shown that, when a cavitation cluster is stabilized near the end face of the cylinder, the latter experiences the action of an attracting force that depends on the cavitation intensity, as well as on the dynamics and shape of the cluster. Eventually, the dependence of the force on the ultrasound frequency has a resonant form.

Under definite conditions, ultrasonically induced cavitation in the gap causes the pendulum to self-oscillate. The self-oscillation frequency is several orders of magnitude lower than the frequency of the forcing ultrasound and may fall into the audible range, causing noise and vibration of a device.

These effects may be useful in designing cavitation sensors and ultrasonic source protection, as well as in turning a system under cavitation to resonance.

ACKNOWLEDGMENTS

The authors thank L.I. Mal’tsev and V.S. Teslenko for support and valuable comments and also S.V. Surodin for the development of the rectangular pulse generator used in this work.

This work was supported by the Russian Foundation for Basic Research (grant no. 03-02-17682); the program “Integration” at the Siberian Division, Russian Academy of Sciences (project no. 123); and the German Academic Exchange Service (DAAD).

REFERENCES

1. G. I. Kuvshinov and P. P. Prokhorenko, *Acoustic Cavitation near Solid Surfaces*, Ed. by V. K. Kedrinskii (Navuka i Tekhnika, Minsk, 1990) [in Russian].
2. P. P. Prokhorenko, N. V. Dezhkunov, and G. E. Konovalov, *The Ultrasonic Capillary Effect* (Nauka i Tekhnika, Minsk, 1981) [in Russian].
3. N. V. Malykh, A. O. Kuz’min, V. M. Petrov, and O. P. Pestunova, in *Proceedings of the 11th Session of the Russian Acoustic Society*, Vol. 2: *Acoustic Measure-*

ment, Geoacoustics, Electroacoustics, and Ultrasonics (GEOS, Moscow, 2001), pp. 296–299.

4. F. Jackson, J. Acoust. Soc. Am. **32**, 1387 (1960).
5. I. Hanson, V. Kedrinskii, and K. A. Merch, J. Phys. D **15**, 1725 (1982).
6. V. K. Kedrinskii, Prikl. Mekh. Tekh. Fiz. **37** (4), 22 (1996).
7. N. V. Malykh, V. M. Petrov, and G. N. Sankin, in *Proceedings of the 5th World Congress on Ultrasonics (WCU-2003), Paris, 2003,* pp. 1343–1346.
8. J. R. Blake, G. S. Keen, R. P. Tong, and M. Wilson, Philos. Trans. R. Soc. London, Ser. A **357**, 251 (1999).
9. N. V. Dezhkunov, A. Francescutto, P. Ciuti, and P. Ignatenko, in *Proceedings of the 5th World Congress on Ultrasonics (WCU-2003), Paris, 2003*, pp. 597–600.
10. N. V. Malykh and V. M. Petrov, in *Proceedings of the 13th Session of the Russian Acoustic Society*, Vol. 1: *Physical Acoustics* (GEOS, Moscow, 2003), pp. 40–43.

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 924–929. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 106–111.
Original Russian Text Copyright © 2005 by Eliseeva, Sementsov.

Electromagnetic Eigenwave Spectrum in a Periodic Ferromagnet–Semiconductor Structure

S. V. Eliseeva and D. I. Sementsov
Ul'yanovsk State University, Ul'yanovsk, 432970 Russia
e-mail: veliseyeva@bk.ru; sdi@sdi.ulsu.ru

Received June 7, 2004; in final form, November 25, 2004

Abstract—The propagation and spectrum of eigenwaves in a periodic ferromagnet–semiconductor structure subjected to a plane bias field perpendicular to the propagation direction are studied. Transformation matrices for the structure period and dispersion relations for the *TE* and *TM* waves are obtained and analyzed in the wavelength region where the spectrum exhibits a band structure. It is demonstrated with the reflection coefficient that the external field can be used to control the eigenwave properties in different ranges. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Electromagnetic wave propagation in multilayer periodic structures consisting of different materials has attracted considerable attention of researchers for many years [1–3]. In terms of parameter control efficiency, semiconductors [4] and magnetics [5, 6] seem to be the most promising for periodic structures, since their characteristics can easily be varied by applying electric or magnetic fields. One of the most important characteristics of a periodic structure is its band spectrum, which relates the frequency and wavevector of eigenwaves propagating in it. There is direct analogy between electromagnetic waves in periodic structures and electron density waves in a periodic crystal field, which follows from translation symmetry and, as a consequence, from similar dispersion relations. Therefore, a number of features, such as forbidden and allowed energy (frequency) bands, are typical of both electron and electromagnetic waves [7].

The band spectrum of semiconductor–insulator periodic structures was comprehensively studied in [8–10]. For helical and multidomain magnetogyrotropic periodic structures, the spectrum was carefully studied in the optical range, where the gyrotropic properties of a magnetic depend on its permittivity [11–13]. For layered ferromagnet–insulator structures, a general dispersion relation at microwave frequencies was studied in the thin-layer approximation, which assumes that the period of the structure is much shorter than the wavelength in the medium ($d \ll \lambda$) [14, 15]. Of particular applied interest may be a magnetic–semiconductor structure, which exhibits the gyrotropic properties for *TE* waves in the microwave range (due to magnetic gyrotropy) and for *TM* waves in the IR range (due to electric gyrotropy). In this paper, we study the eigenwave spectrum in a medium consisting of alternating magnetic insulating and nonmagnetic semiconductor layers subjected to an external plane magnetic field both in the thin-layer approximation and under conditions when the spectrum exhibits a band structure.

GENERAL EQUATIONS

Consider a planar layered periodic structure consisting of magnetic layers with thickness d_1 and semiconductor layers with thickness d_2. The *OZ* axis of the Cartesian system is perpendicular to the interfaces. An external bias field is aligned with the *OX* axis and an electromagnetic wave propagates along the interfaces (along the *OY* axis) normally to the external bias field. The high-frequency properties of magnetic layers are described through their permeability, which is generally a tensor. Magnetics are known to exhibit tensor properties due to magnetic gyrotropy in the microwave range [16]. For the coordinate system and bias field direction chosen, the nonzero components of the permeability of an isotropic magnetic have the frequency dependences

$$\mu_{yy} = \mu_{zz} = 1 + \frac{\omega_M(\omega_H^2 + i\omega_r\omega)}{\omega_H(\omega_H^2 - \omega^2 + 2i\omega_r\omega)},$$

$$\mu_{yz} = -\mu_{zy} = \frac{i\omega\omega_M}{\omega_H^2 - \omega^2 + 2i\omega_r\omega}, \qquad (1)$$

$$\mu_{xx} = 1.$$

Here, $\omega_M = 4\pi\gamma M$, $\omega_H = \gamma H$, and $\omega_r = \xi\omega_H$, M is the saturation magnetization, H is the external static bias field, γ is the gyromagnetic ratio, and ξ is the relaxation parameter of the magnetic system. For electric properties, the ferromagnet also represents an isotropic medium; therefore, its permittivity tensor is a diagonal matrix with components ε_f.

For semiconductor layers, the nonzero components of the permittivity have the form [4]

$$\varepsilon_{yy} = \varepsilon_{zz} = \varepsilon_0 - \frac{\omega_{\mathrm{p}}^2(\omega + i\nu)}{\omega[(\omega+i\nu)^2 - \omega_{\mathrm{c}}^2]},$$

$$\varepsilon_{yz} = -\varepsilon_{zy} = -\frac{i\omega_{\mathrm{p}}^2\omega_{\mathrm{c}}}{\omega[(\omega+i\nu)^2 - \omega_{\mathrm{c}}^2]}, \tag{2}$$

$$\varepsilon_{xx} = \varepsilon_0.$$

Here, ω_{p} is the plasma frequency; $\omega_{\mathrm{c}} = eH/m^*c$ is the cyclotron frequency; m^* is the effective mass of carriers; ν is the effective collision frequency; and ε_0 is the lattice part of the permittivity, which, in general, depends on frequency [17]. The permeability tensor of a nonmagnetic semiconductor will also be considered to be a diagonal matrix with components μ_{s} close to 1.

Solving the Maxwell equations for each of the layers with regard to the propagation direction and material parameters of the layers yields two eigenmodes: the *TE* mode with field components (e_x, h_y, h_z) and the *TM* mode with field components (h_x, e_y, e_z). Let us write expressions for the fields of these differently polarized waves in each of the media. All the wave field components are assumed to vary with time as $\exp(i\omega t)$ and along the propagation direction as $\exp(-iky)$, where k is longitudinal component of the wavevector (i.e., the propagation constant). These exponential factors are further omitted.

The components of the *TE* wave propagating in the magnetic layers depend on the transverse coordinate as follows:

$$e_{x\mathrm{f}} = A_1\exp(i\nu_{\mathrm{f}}z) + A_2\exp(-i\nu_{\mathrm{f}}z),$$

$$h_{y\mathrm{f}} = \frac{i}{k_0\mu_\perp}\frac{de_{x\mathrm{f}}}{dz} + \frac{k\mu_{yz}}{k_0\mu_\perp\mu_{zz}}e_{x\mathrm{f}}, \tag{3}$$

$$h_{z\mathrm{f}} = -\frac{k\mu_{yy}}{k_0\mu_{zz}\mu_\perp}e_{x\mathrm{f}} - \frac{i\mu_{zy}}{k_0\mu_{zz}\mu_\perp}\frac{de_{x\mathrm{f}}}{dz},$$

where $k_0 = \omega/c$, ω and c are the frequency and velocity of the wave in a vacuum, $\nu_{\mathrm{f}} = (k_0^2\varepsilon_{\mathrm{f}}\mu_\perp - k^2)^{1/2}$ is the transverse component of the wavevector, and $\mu_\perp = \mu_{yy} - \mu_{yz}\mu_{zy}/\mu_{zz}$ is the effective permeability of the magnetic layers.

The expression for $\mu_\perp$ in view of (1) implies that the characteristic frequencies of the magnetic layers are ferromagnetic resonance frequency $\omega_{\mathrm{f}} = \sqrt{\omega_H(\omega_H + \omega_M)}$, at which $\mu_\perp \longrightarrow \infty$ if magnetic relaxation is neglected, and antiresonance frequency $\omega_{\mathrm{a}} = \omega_H + \omega_M$, at which $\mu_\perp = 0$. It then follows that, at microwave frequencies, the characteristics of the propagating wave considerably depend on the external magnetic field and, thus, may be effectively controlled by it.

For the semiconductor layers, the field components of the *TE* wave can also be represented by expressions (3) with the substitutions $f \longrightarrow s$, $\mu_\perp = \mu_{\alpha\alpha} \longrightarrow \mu_{\mathrm{s}}$, $\mu_{\alpha\beta} = 0$, and $\varepsilon_{\mathrm{f}} \longrightarrow \varepsilon_0$. The characteristic frequencies of lattice part $\varepsilon_0(\omega)$ of the semiconductor permittivity are several orders of magnitude higher than microwave frequencies. Therefore, in the microwave range, ε_0 can be regarded as a constant and insulating layers, a passive medium for the *TE* wave.

The *TM* wave field in the semiconductor layers can be written as

$$h_{x\mathrm{s}} = B_1\exp(i\nu_{\mathrm{s}}z) + B_2\exp(-i\nu_{\mathrm{s}}z),$$

$$e_{y\mathrm{s}} = -\frac{i}{k_0\varepsilon_\perp}\frac{dh_{x\mathrm{s}}}{dz} - \frac{k\varepsilon_{yz}}{k_0\varepsilon_\perp\varepsilon_{zz}}h_{x\mathrm{s}}, \tag{4}$$

$$e_{z\mathrm{s}} = \frac{k\varepsilon_{yy}}{k_0\varepsilon_{zz}\varepsilon_\perp}h_{x\mathrm{s}} + \frac{i\varepsilon_{zy}}{k_0\varepsilon_{zz}\varepsilon_\perp}\frac{dh_{x\mathrm{s}}}{dz},$$

where $\nu_{\mathrm{s}} = (k_0^2\varepsilon_\perp\mu_{\mathrm{s}} - k^2)^{1/2}$ and $\varepsilon_\perp = \varepsilon_{yy} - \varepsilon_{yz}\varepsilon_{zy}/\varepsilon_{zz}$.

The expression for $\varepsilon_\perp$ in view of (2) shows that the characteristic frequencies of the semiconductor layers are $\omega_{\mathrm{s}} = \sqrt{\omega_{\mathrm{c}}^2 + \omega_{\mathrm{p}}^2/\varepsilon_0}$, at which $\varepsilon_\perp \longrightarrow \infty$ if collisions are neglected, and the frequencies at which $\varepsilon_\perp = 0$; that is,

$$\omega_{\mathrm{b}}^{(\pm)} = \frac{1}{\sqrt{2}}\left[\omega_{\mathrm{c}}^2 + 2\frac{\omega_{\mathrm{p}}^2}{\varepsilon_0} \pm \omega_{\mathrm{c}}\sqrt{\omega_{\mathrm{c}}^2 + 4\frac{\omega_{\mathrm{p}}^2}{\varepsilon_0}}\right]^{1/2}.$$

For parameter values typical of a semiconductor (for example, InSb), $\omega_{\mathrm{p}} = 4.81 \times 10^{12}\ \mathrm{s}^{-1}$, $\omega_{\mathrm{c}} = 3.2 \times 10^{11}\ \mathrm{s}^{-1}$, $\nu = 10^{10}\ \mathrm{s}^{-1}$, and $\varepsilon_0 = 17.8$, we have $\omega_{\mathrm{s}} \cong 1.18 \times 10^{12}\ \mathrm{s}^{-1}$, $\omega_{\mathrm{b}}^{+} \cong 1.29 \times 10^{12}\ \mathrm{s}^{-1}$, and $\omega_{\mathrm{b}}^{-} \cong 0.98 \times 10^{12}\ \mathrm{s}^{-1}$. These frequencies are seen to fall into the far-IR range, where the characteristics of the *TM* wave in a semiconductor can be controlled by applying a biasing field.

In the magnetic layers, the components of the *TE* wave can be expressed by formulas (4) with the substitutions $s \longrightarrow f$, $\varepsilon_\perp = \varepsilon_{\alpha\alpha} \longrightarrow \varepsilon_{\mathrm{f}}$, $\varepsilon_{\alpha\beta} = 0$, and $\mu_{\mathrm{s}} \longrightarrow \mu_{xx} = 1$. For the *TM* wave, the magnetic layers with the bias field direction chosen represent a passive medium that cannot be controlled by an external magnetic field.

TRANSFORMATION MATRIX AND DISPERSION RELATION

To find the amplitude distribution of the wave field in an infinite periodic layered structure, we introduce transformation matrix $\hat{m}$ for two layers constituting the period. This matrix relates the tangential components of the field at the beginning and at the end of the period. In particular, the amplitudes of the *TE* wave field components at the boundaries of nth and $(n-2)$th layers are related as follows:

$$\begin{aligned}\varepsilon_{xn} &= m_{11}e_{xn-2} + m_{12}h_{yn-2},\\ h_{yn} &= m_{21}e_{xn-2} + m_{22}h_{yn-2}.\end{aligned} \tag{5}$$

Taking into account that the tangential field components at the interfaces must be continuous and periodic,

$$e_{x\mathrm{f}}(d_1) = e_{x\mathrm{s}}(d_1), \quad h_{y\mathrm{f}}(d_1) = h_{y\mathrm{s}}(d_1), \tag{6}$$

we arrive at expressions for the transformation matrix elements,

$$m_{11} = C_1C_2 - i\frac{k\mu_{yz}}{\nu_\mathrm{f}\mu_{zz}}S_1C_2 - \frac{\mu_\perp\nu_\mathrm{s}}{\mu_\mathrm{s}\nu_\mathrm{f}}S_1S_2,$$

$$m_{22} = C_1C_2 - i\frac{k\mu_{yz}}{\nu_\mathrm{f}\mu_{zz}}S_1C_2 + \frac{\mu_\mathrm{s}}{\nu_\mathrm{s}\mu_\perp}\left(\frac{k^2\mu_{yz}^2}{\nu_\mathrm{f}\mu_{zz}^2} - \nu_\mathrm{f}\right)S_1S_2,$$

$$m_{12} = -\frac{k_0\mu_\mathrm{s}}{i\nu_\mathrm{s}}C_1S_2 + \frac{k\mu_{yz}}{\nu_\mathrm{f}\mu_{zz}}\frac{k_0\mu_\mathrm{s}}{\nu_\mathrm{s}}S_1S_2 - \frac{k_0\mu_\perp}{i\nu_\mathrm{f}}S_1C_2, \tag{7}$$

$$m_{21} = \frac{i\nu_\mathrm{s}}{k_0\mu_\mathrm{s}}C_1S_2 - \frac{k\mu_{yz}}{\nu_\mathrm{f}\mu_{zz}}\frac{\nu_\mathrm{s}}{k_0\mu_\mathrm{s}}S_1S_2 - i\frac{1}{k_0\mu_\perp}\left(\frac{k^2\mu_{yz}^2}{\nu_\mathrm{f}\mu_{zz}^2} - \nu_\mathrm{f}\right)S_1C_2,$$

where $C_1 = \cos\nu_\mathrm{f}d_1$, $C_2 = \cos\nu_\mathrm{s}d_2$, $S_1 = \sin\nu_\mathrm{f}d_1$, and $S_2 = \sin\nu_\mathrm{s}d_2$.

Proceeding in a similar way, we can derive the components of transformation matrix $\hat{m}$ for the TM wave.

A dispersion relation for electromagnetic waves in a medium consisting of alternating magnetic and semiconductor layers can easily be derived using the periodicity conditions $e_{x\mathrm{s}}(d) = e_{x\mathrm{f}}(0)\exp(i\nu_\mathrm{eff}d)$ and $h_{y\mathrm{s}}(d) = h_{y\mathrm{f}}(0)\exp(i\nu_\mathrm{eff}d)$,

$$2\cos\nu_\mathrm{eff}d = m_{11} + m_{22}, \tag{8}$$

where ν_eff is the effective Bloch wavenumber, which is actually the period-averaged transverse component of the wavevector of the wave propagating in the structure.

With regard to the expressions for the diagonal elements of matrix $\hat{m}$ for the *TE* and *TM* waves, this relationship can be reduced to

$$\begin{aligned}\cos(\nu_\mathrm{eff}^{TE,\,TM}d) &= C_1C_2 - G^{TE,\,TM}S_1S_2,\\ G^{TE} &= \frac{\mu_\mathrm{s}}{2\mu_\perp}\left(\frac{\mu_\perp^2\nu_\mathrm{s}}{\mu_\mathrm{s}^2\nu_\mathrm{f}} + \frac{\nu_\mathrm{f}}{\nu_\mathrm{s}} - \frac{k^2}{\nu_\mathrm{f}\nu_\mathrm{s}}\frac{\mu_{yz}^2}{\mu_{zz}^2}\right),\\ G^{TM} &= \frac{\varepsilon_\mathrm{f}}{2\varepsilon_\perp}\left(\frac{\varepsilon_\perp^2\nu_\mathrm{s}}{\varepsilon_\mathrm{f}^2\nu_\mathrm{f}} + \frac{\nu_\mathrm{f}}{\nu_\mathrm{s}} - \frac{k^2}{\nu_\mathrm{f}\nu_\mathrm{s}}\frac{\varepsilon_{yz}^2}{\varepsilon_{zz}^2}\right).\end{aligned} \tag{9}$$

Dispersion relations (9) specify the spectrum of electromagnetic modes propagating in a periodic structure with a given biasing field configuration and propagation direction.

ANALYSIS OF THE DISPERSION RELATION

Solutions to dispersion relations (9) exist for both real and imaginary values of parameters ν_f and ν_s, i.e., for both bulk and surface collective polariton waves. These relationships have the simplest form in the thin-layer approximation, when the conditions $\nu_\mathrm{f}d_1, \nu_\mathrm{s}d_2 \ll 1$ are satisfied. In this case, the effective Bloch wavenumbers for eigenwaves are given by

$$\begin{aligned}\nu_\mathrm{eff}^{TE} &= \frac{1}{1+\theta}\left[\nu_\mathrm{f}^2\theta^2\left(1 + \frac{\mu_\mathrm{s}}{\theta\mu_\perp}\right) + \nu_\mathrm{s}^2\left(1 + \frac{\theta\mu_\perp}{\mu_\mathrm{s}}\right) - \frac{\mu_{yz}^2\mu_\mathrm{s}}{\mu_{zz}^2\mu_\perp}\theta k^2\right]^{1/2},\\ \nu_\mathrm{eff}^{TM} &= \frac{1}{1+\theta^{-1}}\left[\nu_\mathrm{s}^2\theta^{-2}\left(1 + \frac{\varepsilon_\mathrm{f}}{\theta^{-1}\varepsilon_\perp}\right) + \nu_\mathrm{f}^2\left(1 + \frac{\theta^{-1}\varepsilon_\perp}{\varepsilon_\mathrm{f}}\right) - \frac{\varepsilon_{yz}^2\varepsilon_\mathrm{f}}{\varepsilon_{zz}^2\varepsilon_\perp}\theta^{-1}k^2\right]^{1/2},\end{aligned} \tag{10}$$

where $\theta = d_1/d_2$ and parameters $\theta/(1 + \theta)$ and $1/(1 + \theta)$ determine the contributions of the magnetic and semiconductor layers, respectively, to structure period $d = d_1 + d_2$.

At frequencies $\omega = (10^{10}\text{–}10^{12})\ \mathrm{s}^{-1}$, relationships (10) are valid for structures like that under study at $d < 10^2\ \mu\mathrm{m}$. As follows from expressions (10), the characteristics of a layered medium may be varied over a wide range by varying both a bias field and the semiconductor-to-magnetic layer thickness ratio.

Figure 1 shows the frequency dependences of the real and imaginary parts of the effective Bloch wavenumbers for the *TE* and *TM* modes near the resonance frequency of parameters $\mu_\perp$ and $\varepsilon_\perp$ at bias field $H = 2000$ Oe. The real part of the effective Bloch wavenumber (thick line) specifies the effective wavelength of the periodic part of the field distribution along the z coordinate, i.e., $\lambda_\mathrm{eff} = 2\pi/\mathrm{Re}(\nu_\mathrm{eff}^{TE,\,TM})$. The imaginary part of the Bloch wavenumber, $\nu_\mathrm{eff}^{TE,\,TM}$, which is responsible for the penetration depth of the wave into the periodic structure, $\delta_\mathrm{eff}^{-1} = \mathrm{Im}(\nu_\mathrm{eff}^{TE,\,TM})$, is plotted by the thin line. Hereafter, calculations were made for the following values of the characteristic frequencies and parameters: $\omega_M = 3.11 \times 10^{10}\ \mathrm{s}^{-1}$, $\omega_H = 3.52 \times 10^{10}\ \mathrm{s}^{-1}$, $\omega_\mathrm{r} = 1.06 \times 10^8\ \mathrm{s}^{-1}$, $\varepsilon_\mathrm{f} = 5.5$, and $\theta = 1$. The real and imaginary parts of ν_eff^{TE} reach maxima at ferromagnetic resonance frequency $\omega_f = 4.83 \times 10^{10}$ s^{-1} of the effective permeability. A minimum of the real part of the Bloch wavenumber is observed in the interval $\omega_f < \omega < \omega_a$, where $\omega_a = 6.63 \times 10^{10}$ s^{-1} is the antiferromagnetic resonance frequency. A minimum of the imaginary part of ν_{eff}^{TE} is observed in the interval $\omega_a < \omega < \omega_f$. For the *TM* wave, the real and imaginary parts of ν_{eff}^{TM} reach a maximum at the resonance frequency of effective permittivity $\varepsilon_\perp$ (and not of the effective permeability). A minimum of $\text{Im}(\nu_{\text{eff}}^{TM})$ is observed in the intervals $\omega_b^- < \omega < \omega_s$ and $\omega > \omega_b^+$.

At layer thicknesses $d_1 \geq 10^2$ μm, the thin-layer approximation fails and relationships (9) need a more rigorous analysis. At frequencies $\omega \gg \omega_f$, at which magnetic layers lose gyrotropic properties, a ferromagnet–semiconductor structure is similar to a semiconductor–insulator structure in electromagnetic characteristics. The analysis of general dispersion relation (9) for the *TM* wave in a semiconductor–dielectric structure is given in [8–10]. We will concentrate on general dispersion relation (9) for the *TE* wave as applied to a symmetric structure in the microwave range, neglecting magnetic relaxation ($\omega_r = 0$).

In analyzing dispersion relation (9), dependences $\omega(k)$ given by

$$k_1 = \frac{\omega}{c}\sqrt{\varepsilon_0\mu_s}, \quad k_2 = \frac{\omega}{c}\sqrt{\varepsilon_f\mu_\perp(\omega)}, \tag{11}$$

are of great importance, because, at $k > k_{1,\,2}$, transverse wavenumbers ν_s and ν_f are imaginary and the associated polariton waves are surface waves. These dependences (the frequency against the normalized wavenumber) are plotted in Fig. 2 (curves *1* and *2*). In the hatched and unhatched areas, ν_s and ν_f are real (and so bulk waves) and imaginary (surface waves localized at the interfaces), respectively. In the doubly hatched areas, bulk waves exist in both media; in the singly hatched areas, bulk waves exist in one medium and surface ones in the other. The propagation constant is normalized by $k_f = \omega_f/c$. The characteristic frequencies shown in Fig. 2 considerably depend on the bias field, which makes control of the type and parameters of waves propagating in a magnetogyrotropic structure feasible.

Figure 3 plots the frequency dependence of parameter ν_{eff} calculated by (9) for three values of normalized propagation constant k/k_f = (a) 1.86, (b) 4.97, and (c) 8.69, which correspond to different regions on the (ω, k) diagram. Allowed and forbidden frequency regions are seen, indicating that the spectrum of the collective waves has a band structure. The spectrum is periodic in Bloch wavenumber with a period of $2\pi/d$; therefore, it is shown for each k in that interval of ν_{eff} corresponding to one of the Brillouin zones. The allowed bands, which support propagation of the collective bulk and surface waves, occupy the frequency interval between Bloch wavenumbers $\nu_{\text{eff}} = 2\pi n/d$ and $\pi(2n + 1)/d$, where $n = 0, 1, 2, \ldots$. In the range $\omega < \omega_f$, the allowed bands come closer together, both the allowed and forbidden bands becoming narrower. This is because, as the frequency approaches the ferromagnetic resonance frequency, $\mu_\perp$ tends to infinity and so does the optical thickness of the ferromagnet layers, which causes rapid oscillations of the trigonometric functions in dispersion relationship (9) and generates many pass and stop bands. With an increase in the propagation constant, bands with the same number shift

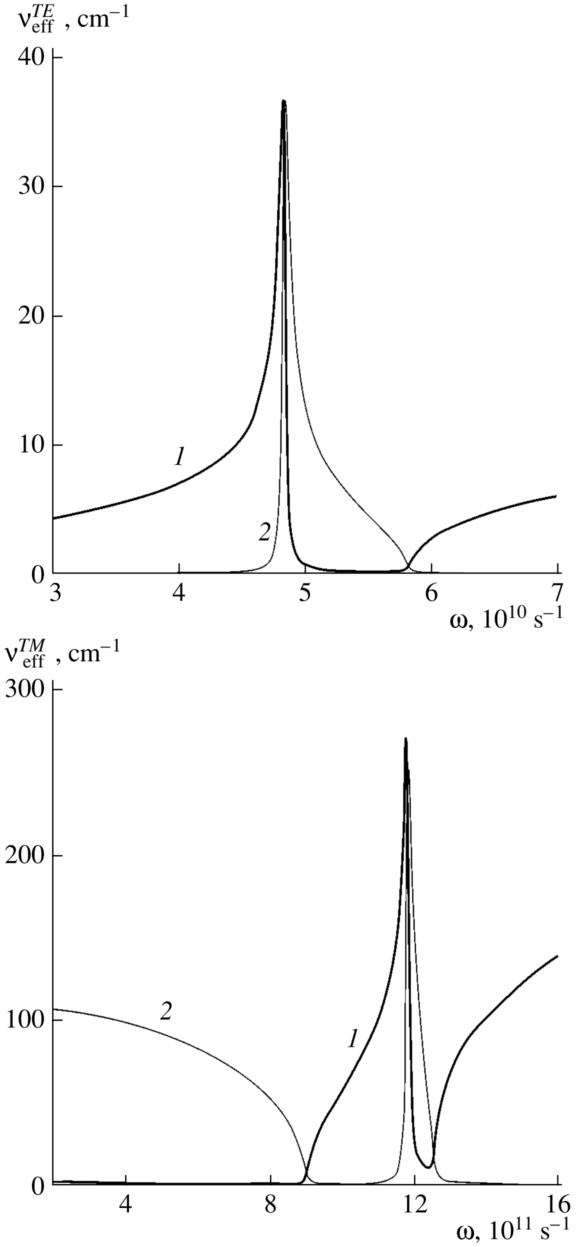


Fig. 1. Effective Bloch wave number ν_{eff} for the (*1*) *TE* and (*2*) *TM* waves vs. frequency.

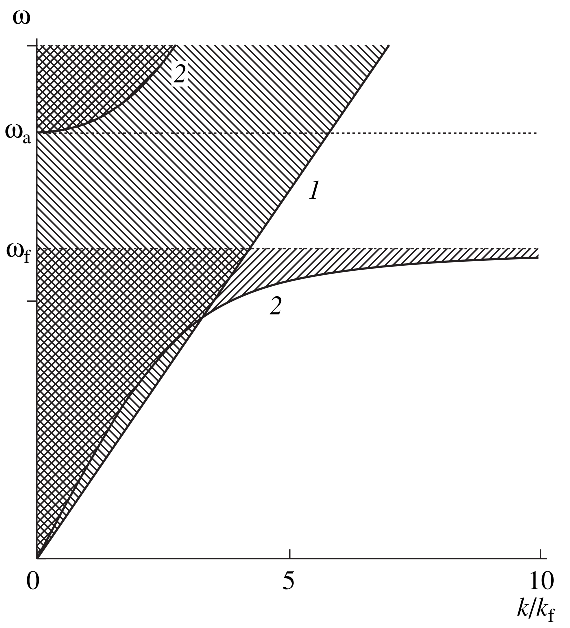


Fig. 2. Frequency vs. propagation constant diagram defining the ranges of existence of the bulk and surface waves. Curves *1* and *2* are described by Eqs. (11).

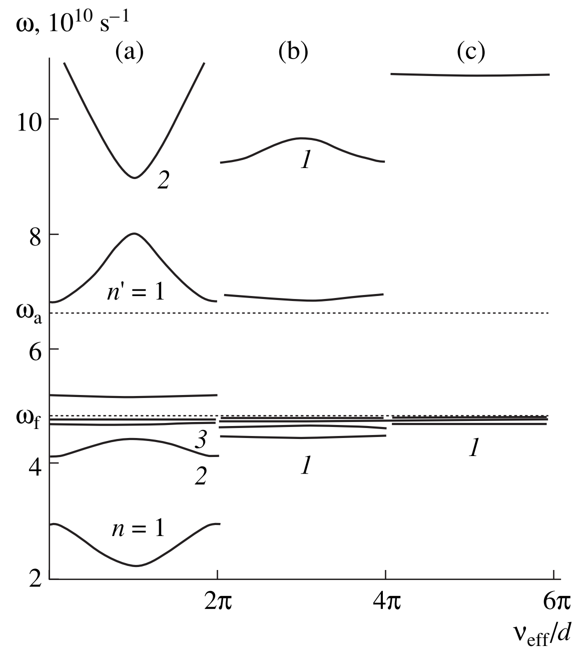


Fig. 3. Frequency vs. normalized Bloch wavenumber of the periodic structure at k/k_f = (a) 1.86, (b) 4.97, and (c) 8.69.

toward higher frequencies. At $\omega > \omega_f$, a single narrow band appears, corresponding to a surface polariton mode. At small k, this mode is akin to a surface wave on the side of the magnetic layers and to a bulk wave on the side of the semiconductor layers. At $k > k_1(\omega)$, this mode becomes a surface wave on the side of the semiconductor layers as well. At $\omega > \omega_a$, bulk-wave bands appear again. Their numbers are $n' = 1, 2, 3, \ldots$.

REFLECTION COEFFICIENT SPECTRUM

A deeper insight into the eigenwave spectrum in the periodic structure under study can be gained with experiments on reflection of electromagnetic waves incident on the structure from a homogeneous medium. Prior to experiments, however, it would be reasonable to define the coefficient of reflection of the TE wave from our periodic structure and analyze its behavior versus the parameters of the structure and incident wave. Let a plane wave with frequency ω and wavenumber $k_i = k_0\sqrt{\varepsilon\mu}$ be incident from the domain $z < 0$ occupied by a homogeneous nonmagnetic medium of permittivity ε and permeability μ on a layered periodic structure occupying the domain $z > 0$ normally to the interfaces. In this case, the total wave field in the domain $z < 0$ is a superposition of the incident and reflected fields,

$$\begin{aligned} h_y &= h_y^{(i)}[\exp(ik_iz) + r\exp(-ik_iz)], \\ e_x &= -\sqrt{\mu/\varepsilon}\,h_y^{(i)}[\exp(ik_iz) - r\exp(-ik_iz)], \end{aligned} \qquad (12)$$

where $r = h_y^{(r)}/h_y^{(i)}$ is the complex amplitude reflection coefficient and $h_y^{(r)}$ and $h_y^{(i)}$ are the amplitudes of the reflected and incident waves.

The reflection coefficient will be sought using the expressions for the field in either of the media, boundary conditions stated above, periodicity conditions (6), and boundary conditions

$$e_x(0) = e_{xf}(0), \quad h_y(0) = h_{yf}(0) \qquad (13)$$

at the interface between the homogeneous half-space and ferromagnet.

Solving the resulting set of equations yields an expression for energy reflection coefficient $R = |r|^2$,

$$R = \left|\frac{\exp(i\nu_{\mathrm{eff}}d) - m_{22} - m_{12}\sqrt{\varepsilon/\mu}}{\exp(i\nu_{\mathrm{eff}}d) - m_{22} + m_{12}\sqrt{\varepsilon/\mu}}\right|^2, \qquad (14)$$

where the longitudinal component of the wavenumber in elements m_{12} and m_{22} of the transmission matrix should be set equal to zero ($k = 0$), because the wave is incident normally to the structure.

Figure 4 plots the reflection coefficient calculated by formulas (14) versus frequency. In the forbidden bands, $R = 1$. In the bands allowed for the bulk waves in the ferromagnet, there are frequencies at which $R \longrightarrow 0$ and the incident wave energy is almost totally transferred to the layered medium. With relaxation neglected, the transferred portion of the energy is specified by transmission coefficient $T = 1 - R$. Throughout the frequency range $\omega_f < \omega < \omega_a$, except for a narrow interval corresponding to the surface polariton mode

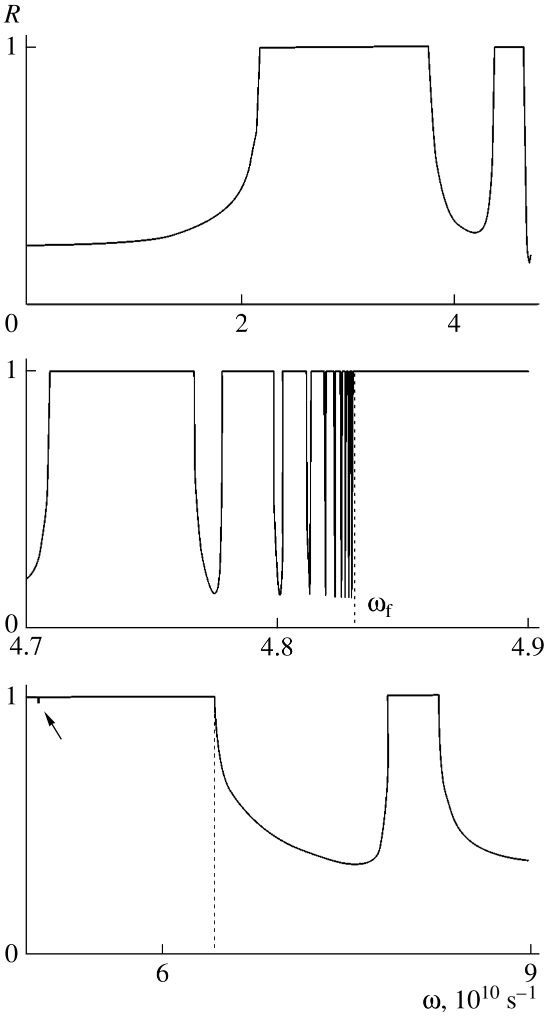


Fig. 4. Reflection coefficient vs. frequency at normal incidence of the wave on the periodic structure ($k = 0$).

(arrow in Fig. 4), total internal reflection is observed ($R = 1$). The small difference of the reflection coefficient from unity in the range where the collective surface mode is excited indicates that this mode is difficult to detect at normal incidence of the wave on the structure.

CONCLUSIONS

We studied the dispersion properties of a ferromagnet–semiconductor structure and conditions under which bulk and surface polariton waves exist. A specific feature of the structure is the feasibility of effectively controlling the parameters of two types of waves of different polarizations in nonoverlapping frequency bands. Varying the thickness ratio of the layers that constitute the period of the structure, one can change the width and number of the pass and stop bands, as well as the depth of field penetration into the structure. A change in an external bias field shifts the characteristic frequencies and also allowed and forbidden frequency bands in the spectrum of collective bulk and surface waves.

REFERENCES

1. L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Nauka, Moscow, 1973; Academic, New York, 1960).
2. Sh. Elashi, Proc. IEEE **64**, 1666 (1976) [Tr. Inst. Inzh. Élektr. Radiotekhn. (TIIER) **64** (12), 22 (1976)].
3. V. A. Belyakov and A. S. Sonin, *Optics of Cholesterics* (Nauka, Moscow, 1982) [in Russian].
4. F. G. Bass, A. A. Bulgakov, and A. P. Tetervov, *High-Frequency Properties of Semiconductors with Superlattices* (Nauka, Moscow, 1989) [in Russian].
5. A. Yelon, in *Physics of Thin Films*, Ed. by H. Francombe and R. W. Hoffman (Academic, New York, 1971; Mir, Moscow, 1973), Vol. 6.
6. A. K. Zvezdin and V. A. Kotov, *Magnetooptics of Thin Films* (Nauka, Moscow, 1988) [in Russian].
7. A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, New York, 1984; Mir, Moscow, 1987).
8. A. A. Bulgakov and O. V. Shramkova, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 712 (2000) [Semiconductors **34**, 686 (2000)].
9. A. A. Bulgakov and O. V. Shramkova, Radiotekh. Élektron. (Moscow) **46**, 236 (2001).
10. A. A. Bulgakov and O. V. Shramkova, Zh. Tekh. Fiz. **73** (3), 87 (2003) [Tech. Phys. **48**, 361 (2003)].
11. M. Sh. Erukhimov and V. V. Tyurnev, Fiz. Tverd. Tela (Leningrad) **17**, 2420 (1975) [Sov. Phys. Solid State **17**, 1618 (1975)].
12. D. I. Sementsov and A. M. Morozov, Fiz. Tverd. Tela (Leningrad) **20**, 2591 (1978) [Sov. Phys. Solid State **20**, 1498 (1978)].
13. D. I. Sementsov, Opt. Spektrosk. **50**, 37 (1981) [Opt. Spectrosc. **50**, 19 (1981)].
14. D. I. Sementsov and G. S. Kosakov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **18**, 1189 (1975).
15. S. V. Eliseeva and D. I. Sementsov, Fiz. Volnovykh Protsessov i Radiotekh. Sistemy **5** (2), 35 (2002); **6** (3), 19 (2003).
16. A. G. Gurevich and G. A. Melkov, *Magnetic Oscillations and Waves* (Nauka, Moscow, 1994; CRC, Boca Raton, 1996).
17. M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *Theory of Waves* (Nauka, Moscow, 1979) [in Russian].

Translated by A. Khzmalyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 930–936. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 112–118.
Original Russian Text Copyright © 2005 by Masunov, Polozov.

ELECTRON AND ION BEAMS, ACCELERATORS

Acceleration and Focusing of Intense Ion Beams in RF Undulator Structures

E. S. Masunov and S. M. Polozov

Moscow Engineering Physics Institute (State University),
Kashirskoe sh. 31, Moscow, 115409 Russia
e-mail: masunov@dinus.mephi.ru

Received June 28, 2004

Abstract—Bunching, acceleration, and transverse focusing of intense ion beams in an undulator linac are considered. Such an accelerator features the absence of an rf field harmonic synchronous with the beam. A 3D equation of motion in the Hamiltonian form is derived in the smooth approximation, and the general conditions for ion beam acceleration and transverse focusing in the undulator linac are formulated. Basic analytical results are compared with the results of numerical simulation of the beam dynamics in the polyharmonic field of an accelerating cavity. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

A challenging task of contemporary acceleration physics is designing low-energy linacs of high-intensity ion beams. These devices have found increasing favor in scientific research and can be applied to advantage, e.g., in power engineering. A bunching accelerator intended for formation and acceleration of light ions with energies varying in the ranges from 50–150 keV to 1–3 MeV, beam currents from several tens of milliamperes to several hundreds of milliamperes, and a current transmission coefficient close to unity is the most difficult to develop. In acceleration of intense ion beams, the self-field of the beam's space charge has a considerable effect, especially at low energies, and provision of efficient transverse focusing is a bottleneck. For slow ions, this problem can hardly be solved with external focusing elements, and so both longitudinal and transverse focusings must be accomplished by tailoring the field configuration in the system. Today, bunching accelerators using an rf quadrupole (RFQ) [1] and other rf focusing (RFF) techniques [2, 3] are the most popular. In the RFQ accelerators designed in Russia and abroad, beam currents of 100–150 mA (the maximal values for rf accelerators) have been achieved.

Further elevation in the beam current in conventional rf accelerators is a great challenge. To do this requires increasing either the ultimate beam current density (which is virtually infeasible) or the beam diameter (which, in turn, requires widening of the channel aperture). Ribbon beam accelerators are viewed as an alternative to the conventional systems. A high ratio between the transverse sizes in ribbon beams allows for a decrease in the current density by increasing the cross-sectional area without changing the beam intensity, which facilitates transport of a high-current beam in the accelerator channel.

A number of ribbon-beam RFF accelerators were suggested earlier (see, e.g., [4, 5]). However, they suffer from considerable drawbacks, such as a low ultimate current density and the need for a very strong rf field (300 kV/cm). Another version of ribbon beam accelerator is the undulator linac [6–8]. Here, acceleration and focusing of the beam are provided in the field of two electromagnetic waves nonsynchronized with the beam (i.e., in the field of two undulators). Linacs with an electrostatic undulator (UNDULAC-E) and with an rf undulator (UNDULAC-RF, Fig. 1) have tuned out to be the most appropriate for accelerating ribbon beams. In the former case, the beam is accelerated and focused with a spatial harmonic of the rf field and one spatial harmonic of the electrostatic field of the undulator. In

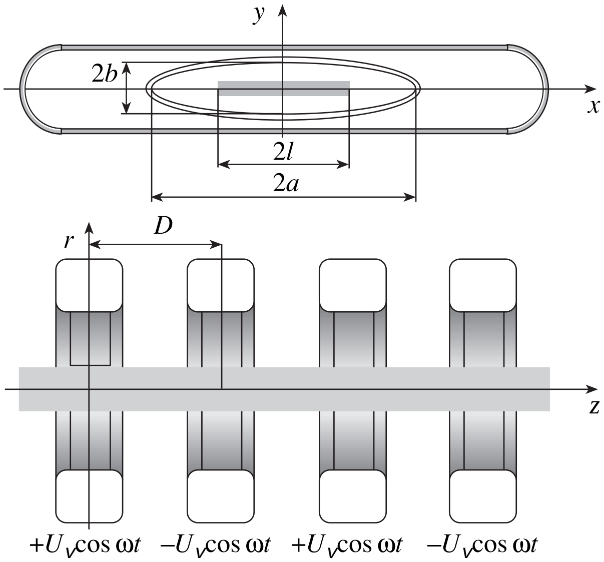


Fig. 1. Design of the linac with an rf undulator.

1063-7842/05/5007-0930$26.00 © 2005 Pleiades Publishing, Inc.

the latter, the beam is accelerated and focused with two or more spatial harmonics of the rf field of the periodic cavity that are nonsynchronized with the beam.

In this work, we carry out detailed analysis of the 3D dynamics of a ribbon beam in an rf undulator accelerator to see whether high-current beam acceleration with this new-type accelerator is feasible.

EQUATION OF MOTION IN THE SMOOTH APPROXIMATION

Comprehensive analysis of the beam dynamics in a polyharmonic field is a challenge. Fast oscillations and the strong dependence of the rf field components on the transverse coordinates that are observed in a ribbon beam accelerators give no way to apply the standard linear approximation to the field expansion in the axial region. For this reason, elaboration of analytical approaches to studying the dynamics seems topical. In [6, 7], it was suggested that the particle dynamics in rapidly oscillating fields be studied with the averaging method (the so-called smooth approximation) and the conditions of applicability of the smooth approximation to the given electrodynamic problem were formulated. As in [7], the rf field in a periodic resonant structure is represented as an expansion in spatial harmonics under the assumption that the period of the structure is a slowly varying function of the longitudinal coordinate. In the quasi-static approximation, the potential of the rf field in the slot channel can be expressed as

$$U = \sum_{n=0}^{\infty} U_n(x, y)\cos\left(\int h_n dz + \alpha\right)\cos(\omega t), \qquad (1)$$

where $h_n = (\mu + 2\pi n)/D$ is the longitudinal wavenumber for an nth rf field harmonic, μ is the oscillation mode and D is the period of the structure.

Function $U_{\mathbf{n}}(x, y)$ satisfied the equation

$$\Delta_\perp U_n = h_n^2 U_n \qquad (2)$$

($\Delta_\perp$ is the transverse component of the Laplacian) and determines the dependence of the potential of the transverse coordinates. Here, two types of solutions are possible. If $U_n(x, y) \sim \cosh(h_{n,x}x)\cosh(h_{n,y}y)$, only the longitudinal component of the electric field strength is present at the axis of the accelerator channel (longitudinal undulator, $\alpha = 0$). If $U_n(x, y) \sim \cosh(h_{n,x}x)\sinh(h_{n,y}y)$, only the transverse components of the electric field are nonzero at the channel axis (transverse undulator, $\alpha = \pi/2$). Here, $h_{n,x}$ and $h_{n,y}$ are the transverse wavenumbers, $h_{n,x}^2 + h_{n,y}^2 = h_n^2$. The ratio $h_{n,x}/h_{n,y}$ specifies the cross-sectional shape of the accelerator channel. The rf field component versus transverse coordinate dependences are easy to find using the relationship $\mathbf{E}_n = -\nabla U_n$.

We will assume that beam velocity β differs markedly from the phase velocity of all field harmonics, $\beta_n = \omega/\cosh_n$ ($n = 0.\ 1, \ldots$) but is close to the velocity of the combined wave that is a superposition of an nth and pth harmonic, $\beta \approx \beta_v = 2\omega/c(h_p \pm h_n)$. Then, a solution to the equation of motion is convenient to seek as a superposition of slowly varying and rapidly oscillating functions. Supposing that amplitude $\tilde{v}$ of fast velocity oscillations is much lower than amplitude v of the slowly varying component, we can write the equation of motion in the smooth approximation in the same way as in [6, 7],

$$\frac{d^2\mathbf{R}}{d\tau^2} = -\nabla U_{\text{eff}}, \qquad (3)$$

where U_{eff} has the meaning of the effective potential function describing the interaction of the particles with the polyharmonic field of the cavity. This function depends only on slowly varying transverse coordinates $\mathbf{R}_\perp = (\rho, \eta)$ (where $\rho = 2\pi x/\lambda$ and $\eta = 2\pi y/\lambda$) and slowly varying phase of the particle in the field of the combined wave $\varphi_x = \int h_v dz - \omega t$, where $h_v = (h_p \pm h_n)/2$ is the longitudinal wavenumber of the combined wave arising as a result of superposition of the fields of the nth and pth harmonics. Now let us introduce the dimensionless amplitudes $e_i = \mathbf{e}\lambda\mathbf{E}_i/2\pi W_0$ of the rf field harmonics, dimensionless time $\tau = \omega t$, and dimensionless longitudinal coordinate $\xi = 2\pi z/\lambda$ (λ and ω are, respectively, the wavelength and frequency of the rf field; W_0 is the energy of the ion being accelerated; and e is the charge of the particle). Then, the effective potential function has the form [7]

$$U_{\text{eff}} = U_1 + U_2 + U_3. \qquad (4)$$

The term

$$U_1(\rho, \eta) = \frac{1}{16}\sum_n \mathbf{e}_n^2\left((\Delta_{n,v}^-)^{-2} + (\Delta_{n,v}^+)^{-2}\right)$$

in Eq. (4) is responsible for transverse focusing, and the terms

$$U_2 = \frac{1}{16}\left(\sum_{h_p - h_n = 2h_v} \mathbf{e}_n\mathbf{e}_p(\Delta_{p,v}^-)^{-2} + \sum_{h_n - h_p = 2h_v} \mathbf{e}_n\mathbf{e}_p(\Delta_{n,v}^-)^{-2}\right)\cos(2\varphi_v)$$

and

$$U_3 = \frac{1}{16}\sum_{h_p + h_n = 2h_v}(e_{z,n}e_{z,p} - \mathbf{e}_{\perp,n}\mathbf{e}_{\perp,p})(\Delta_{n,v}^-)^2 \times \cos(2\varphi_v + 2\alpha_n)$$

are responsible for acceleration of the particles in the longitudinal direction and transverse defocusing of the beam. Here, $\Delta^{\pm}_{n,\,\nu} = (h_n \pm h_\nu)/h_\nu$.

From the above expressions for the effective potential function, one can find the Hamiltonians of the beam–combined wave system in the smooth approximation,

$$\frac{1}{2}\left(\frac{d\mathbf{R}}{d\tau}\right)^2 + U_{\text{eff}} = H. \tag{5}$$

The equation of motion written in form (3) allows a simple analysis of both the longitudinal and transverse motions of particles. Specifically, one can find transverse focusing conditions, the frequencies of transverse and phase oscillations of beam particles, and a relationship between the longitudinal and transverse motions. An intriguing feature of the linac based on an rf undulator is worth noting: the modulation period of the beam is twice as short as the rf field period.

ION BEAM DYNAMICS IN A LONGITUDINAL-FIELD UNDULATOR LINAC

As was mentioned above, the smooth approximation applies if the fast oscillation amplitude is much smaller than a slowly varying velocity of the particles. For low-energy ion beams, this condition is always fulfilled. The beam can be effectively accelerated if the slowly varying velocity of the particles is close to the phase velocity of the combined wave but differs considerably from the phase velocity of the nearest rf field harmonic.

Consider the simplest case when only two spatial rf field harmonics are present in the system ($n = 0$ and $n = 1$). It is reasonable to introduce the notion of the equilibrium particle whose averaged velocity β_x equals β_ν. For oscillations with $\mu = \pi$ mode, the longitudinal wavenumbers are $h_0 = \pi/D$, $h_1 = 3\pi/D$, and $h_s = 2\pi/D$; the velocity is $\beta_s = D/\lambda$; and, in the coordinate system moving with velocity β_ν, U_{eff} can be written as

$$\begin{aligned} U_{\text{eff}} = {} & 1/4\{10/9\mathbf{e}_0^2 + 26/25\mathbf{e}_1^2 + 2(e_{0z}e_{1z} - \mathbf{e}_{0\perp}\mathbf{e}_{1\perp}) \\ & \times[\sin(2\varphi_s + 2\psi + 2\alpha) + 2\psi\cos(2\varphi_s)]\}. \end{aligned} \tag{6}$$

Here, φ_s is the phase of the equilibrium particle and $\psi = \varphi - \varphi_s$. For oscillations with $\mu = 0$ in a cavity with a longitudinal rf field, the wavenumbers of the fundamental and first harmonics and the wavenumber of the combined wave are, respectively, $h_0 = 0$, $h_1 = 2\pi/D$, and $h_v = \pi/D$. Then, $\beta_v = 2D/\lambda$. Accordingly, the effective potential function can be written in the form

$$\begin{aligned} U_{\text{eff}} = {} & 1/8\{\mathbf{e}_0^2 + 5/9\mathbf{e}_1^2 + \mathbf{e}_0\mathbf{e}_1[\cos(2\varphi_s + 2\psi) \\ & + 2\psi\sin(2\varphi_s)] + (e_{0z}e_{1z} - \mathbf{e}_{0\perp}\mathbf{e}_{1\perp}) \\ & \times[\cos(2\varphi_s + 2\psi + 2\alpha) + 2\psi\sin(2\varphi_s)]\}. \end{aligned} \tag{7}$$

For transverse-field undulators, U_{eff} can be found in a similar way.

Using expressions (6) and (7) for the potential function, we arrive at an equation for the velocity increment for near-axis particles in the field of the combined wave,

$$\frac{d\beta}{d\tau} = e_{\text{eff}}\sin 2\varphi, \tag{8}$$

where $e_{\text{eff}} = \nu e_0 e_1/\beta_s$ is the effective amplitude in the field of the combined wave and $\nu = 1$ (if $\mu = \pi$) or 1/2 (if $\mu = 0$).

Consider first the dynamics in a longitudinal-field undulator linac for oscillations with $\mu = \pi$. We assume that amplitudes e_0 and e_1 of the rf field harmonics, as well as phase φ_s of the equilibrium particle, depend on the longitudinal coordinate. To provide effective bunching and acceleration of the beam, the undulator linac must consist of two, bunching and accelerating, sections. For simplicity, phase φ_s of the equilibrium particle in the field of the combined wave in the bunching section is taken to be linearly decreasing from $\pi/2$ to $3\pi/8$ and the amplitudes of the rf field harmonics, increasing by a sine law from initial value $E^{\varepsilon}_{0,\,1} = E_{0,\,1}(z = 0)$. In the accelerating section, φ_s, E_0, and E_1 are constant. In the field of the combined wave, the velocity of the equilibrium particle varies along the accelerator axis according to the equation

$$\frac{d\beta_s^2}{d\tau} = 2e_0(\xi)e_1(\xi)\sin 2\varphi_s(\xi). \tag{8a}$$

Solving this equation, one easily finds the laws of variation of period $D = \beta_s\lambda$ of the structure, as well as of the phase velocities of the fundamental, $\beta_{0,\,s} = 2\beta_s$, and first, $\beta_{1,\,s} = 2\beta_s/3$, rf field harmonics.

Figure 2 plots phase φ_s of the equilibrium particle, amplitudes E_0 and E_1 of the rf field harmonics, and amplitude E_{eff} of the combined wave versus the longitudinal coordinate. The calculation was carried out for a deuterium ion beam with $E_0 = 150$ kV/cm. The other parameters were the following: $\chi = E_1/E_0 = 0.3$, initial ion energy $W_{\text{in}} = 150$ keV ($\beta_{s,\,\text{in}} = 0.013$), accelerator length $L = 2.5$ m, and bunching section length $L_{\text{bn}} = 1$ m.

If the beam velocity were close to the phase velocity of the fundamental (zeroth) harmonic, the synchronous phase in the field of this harmonic would be equal to $\varphi_{0,\,s}$. In the other limiting case (if the beam velocity were close to $\beta_{1,\,s}$), the synchronous phase of the particle would be equal to $\varphi_{1,\,s}$. As follows from Fig. 2, quantities $\varphi_{0,\,s}$ and $\varphi_{1,\,s}$ vary with z only slightly.

From Eq. (8), one can find the separatrix for near-axis particles and, thus, determine the limiting velocity of the particles that can be accelerated,

$$\begin{aligned} (\beta - \beta_s)^2 = {} & e_0 e_1\nu(-(\cos 2\varphi + \cos 2\varphi_s) \\ & + (m\pi - 2\varphi_s - 2\varphi)\sin 2\varphi_s). \end{aligned} \tag{9}$$

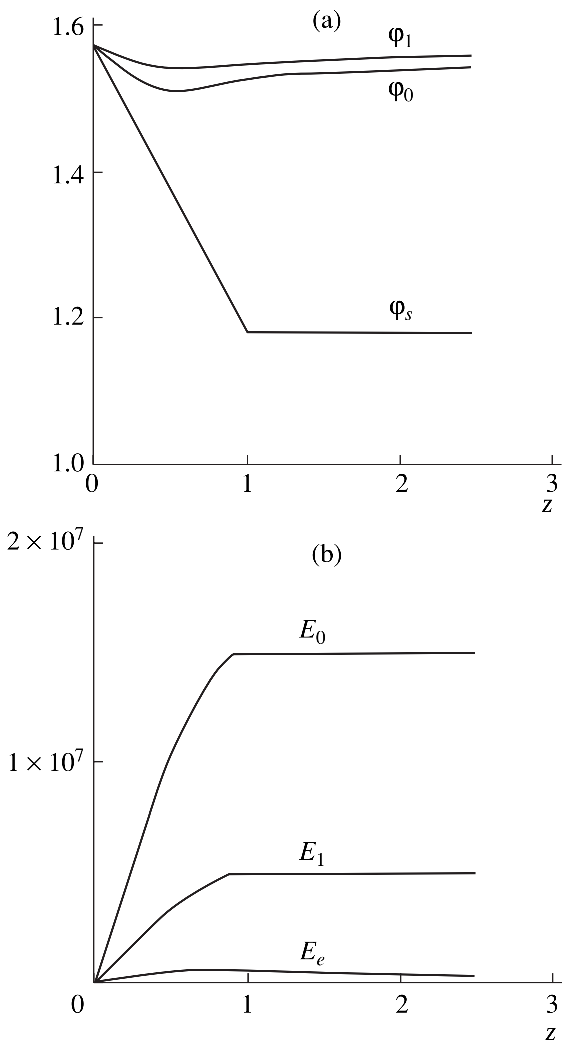


Fig. 2. Variation of the (a) equilibrium particle phase and synchronous phases of the rf field harmonics and (b) amplitudes of the rf field harmonics and combined wave along the *z* direction.

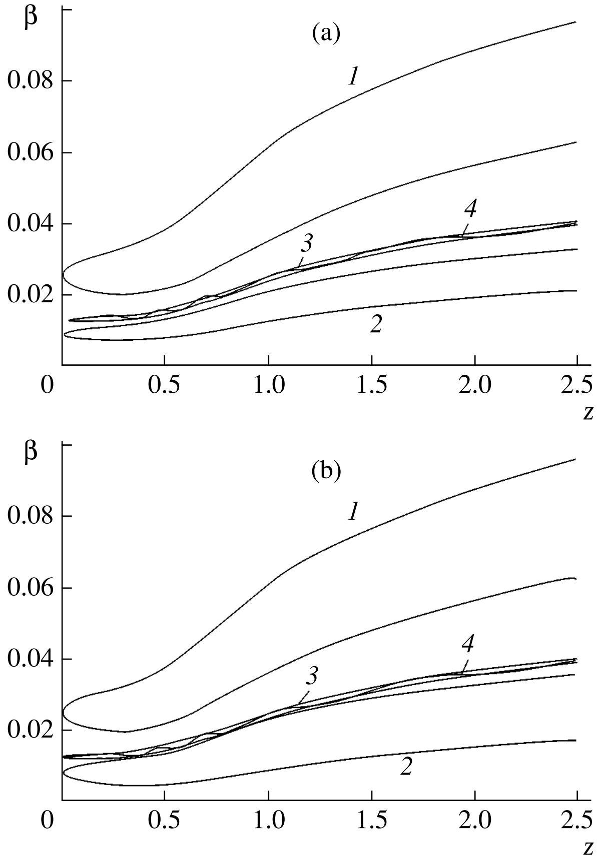


Fig. 3. Variation of the vertical sizes of the separatrices along the *z* direction.

As applied to the combined wave, two phase ranges should be considered: $-\pi < \varphi < 0$ and $0 < \varphi < \pi$. In the range $0 < \varphi < \pi$, $m = 1$; in the range $-\pi < \varphi < 0$, $m = 3$.

As the amplitude of the combined wave grows and φ_s decreases, the vertical (in β) and horizontal (in φ) sizes of the separatrix change. Figure 3 shows the variation of the maximal vertical size of the separatrix of the combined wave with longitudinal coordinate z (curve *3*) for χ = (a) 0.3 and (b) 0.6 in the smooth approximation. Shown also are the same curves obtained for the case when the beam interacts with the zeroth harmonic alone, i.e., under the assumption that velocity β is close to $\beta_{0,s}$ (curve *1*) or $\beta_{1,s}$ (curve *2*). Curve *4* in Fig. 3 shows the z dependence of the longitudinal particle velocity in the smooth approximation. Remarkably, in this approximation, the trajectories of all particles being accelerated are always within the separatrix of the combined wave. The particles can be lost (escape from the acceleration process) only if the adiabatic conditions are disturbed because of a rapid change in the phase velocity and amplitude of the combined wave during bunching. Actually, however, particle losses may be due to other reasons. From Fig. 3, it follows that, when the first harmonic is high (at E_0 = 150 kV/cm, χ must exceed 1/3), the separatrices of the combined wave and rf field first harmonic may partially or completely overlap in the middle of the bunching section. For a small fundamental-to-first harmonic amplitude ratio ($\chi < 0.3$), fast oscillations of the particle longitudinal velocity may far exceed the vertical size of the separatrix (Fig. 4, curves *1*–*3*). Figure 4 shows the longitudinal velocity of the particles with regard to fast oscillations (curve *4*) and the result of its averaging (curve *5*). Even if the separatrices do not overlap, the phase trajectory of the particle may fall within the separatrix of the first harmonic at a certain time. Because of the two effects described above, the particle velocity at some z may approach the phase velocity of the first field harmonic. Strictly speaking, the averaging method fails in this case. Interestingly, when this happens, the particle may be “captured” by the first harmonic in the sense that it is accelerated in its field or even escape the acceleration process. Since these effects are observed at different χ, one may expect that there is an optimal har-

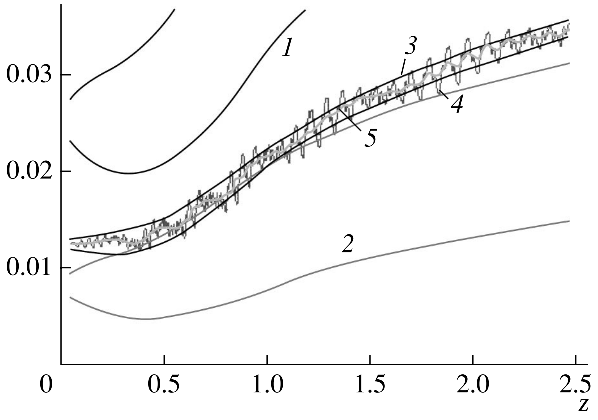


Fig. 4. Effect of the rf field first harmonic on the beam dynamics.

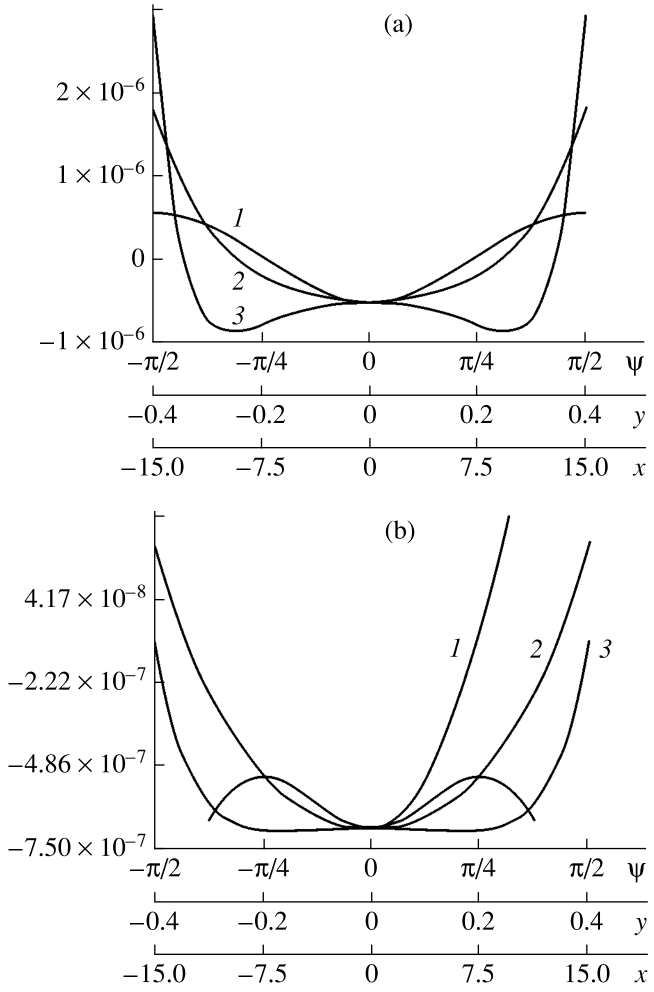


Fig. 5. Sections of the effective potential function.

monic amplitude ratio at which accelerated ion losses are minimal.

A more careful comparison of the dynamics results obtained by calculation in the smooth approximation and by rigorous numerical simulation in the actual field with subsequent averaging showed that, although the amplitudes of slow phase oscillations in these two cases are close to each other, their period obtained rigorously slightly exceeds that calculated in the smooth approximation. Accordingly, the total phase advance of the longitudinal oscillations in the smooth approximation differs from the actual value. The longer the accelerator, the greater the difference between the phase advances. As this difference approaches some critical value, the particle rapidly escapes the acceleration process.

In a similar way, one can analyze the beam dynamics and phase motion of the particles in the rf field for oscillations with $\mu = 0$. In this case, the phase velocities of the combined wave and rf field harmonics are $\beta_s = 2D/\lambda$, $\beta_{0,\,s} = \infty$, and $\beta_{1,\,2} = \beta_s/2$. As in the case of oscillations with $\mu = \pi$, the amplitude of fast oscillations of the particle velocity far exceeds the size of the separatrix calculated in the smooth approximation when χ is small. Here, the velocity does not reach the separatrix of the first harmonic, since the difference between the phase velocities of the combined wave and first harmonic is larger than in the case $\mu = \pi$. At large χ, the separatrices overlap, although the oscillation amplitude of the particle longitudinal velocity is relatively low. Therefore, one can assume that the effect of the first rf harmonic on the longitudinal motion is less pronounced than in the field oscillating with $\mu = \pi$ for the same χ.

The above expression for the 3D effective potential function allows us to find, in the smooth approximation, the condition for transverse stability of the particles in the undulator linac. Indeed, in the absolute minimum of U_{eff}, the conditions for longitudinal and transverse stability of the beam are established simultaneously. In the simple axial approximation ($h_x x \ll 1$ and $h_y y \ll 1$), the transverse focusing condition can be derived analytically. For an undulator linac with field oscillations $\mu = \pi$, it is easy to show that the absolute minimum exists at any harmonic amplitude ratio χ. At the same time, for an undulator linac with field oscillations $\mu = 0$, axial focusing takes place only at $\chi \geq 2$. This is because the rf field first harmonic alone is responsible for transverse focusing of the particles in the latter case. If the dynamics of the beam with large transverse sizes is considered in the nonlinear approximation, transverse focusing is also attainable at smaller χ.

For an undulator linac with field oscillations $\mu = \pi$, Fig. 5 shows sections $U_{\text{eff}}(0, 0, \psi)$ for $x = 0$ and $y = 0$ (curve *1*), $U_{\text{eff}}(0, y, 0)$ for $x = 0$ and $\psi = 0$ (curve *2*), and $U_{\text{eff}}(x, 0, 0)$ for $y = 0$ and $\psi = 0$ (curve *3*). The lots are constructed for $E_0 = 150$ kV/cm; $\chi = 0.6$; $h_x/h_y = 1/23$; and (a) $\varphi_s = \pi/2$, $\beta_s = 0.013$ (the bunching section input) and (b) $\varphi_s = 3\pi/8$, $\beta_s = 0.018$ (the end of the buncher). In the buncher input, U_{eff} has two minima and one maximum along the x axis. As the particle velocity grows, these extrema smooth out and U_{eff} takes an absolute minimum. The presence of the intermediate maximum does not lead to a considerable redistribution of the particle density in the beam, since the transverse oscillations of the particles (across the width of the ribbon beam) have a low frequency: less than one oscillation for a 2.5-m-long accelerator.

More detailed examinations showed that transverse focusing in a longitudinal-undulator linac with $\mu = 0$ is less effective than in the case of $\mu = \pi$ and, therefore, is of no practical interest.

ION BEAM DYNAMICS IN THE TRANSVERSE-FIELD UNDULATOR LINAC

A distinctive feature of the linac with a transverse rf undulator is that here longitudinal acceleration takes place in the presence of only transverse rf fields at the axis of the accelerator. In this case, one cannot consider the beam dynamics only in the axial region. Effective potential U_{eff} is similar in shape to that obtained previously. In the four-dimensional phase space (the motion of the particles normally to the plane of the ribbon is neglected), the Hamiltonian relates the longitudinal and transverse coordinates and velocities,

$$\begin{aligned}(\beta - \beta_s)^2 + \beta_y^2 &= \nu/4(k_0\mathbf{e}_0^2 + k_1\mathbf{e}_1^2) \\ &+ \mathbf{e}_0\mathbf{e}_1 V(2\cosh(2h_y y)\cos(2\varphi_s + 2\psi) + 2\psi\sin(2\varphi_s)).\end{aligned} \tag{10}$$

Here, $k_0 = 1$ and $k_1 = 5/9$ for the oscillations with $\mu = 0$ and $k_0 = 10/9$ and $k_1 = 26/25$ for the oscillations with $\mu = \pi$. Assuming that $\beta_y \ll |\beta - \beta_s|$, one can easily find the projection of the four-dimensional phase volume onto the phase plane (β, φ) and, thereby, determine the condition under which the particles are involved in the acceleration process. Estimates show that the amplitudes of fast oscillations of the longitudinal velocity and phase are small compared with the same parameters in the longitudinal undulator. In the rf undulator linac, the smooth approximation suffices to consider the longitudinal motion. As follows from calculations, the current transmission coefficient for both types of oscillations is 85–90% in the smooth approximation. Here, particle losses are associated only with a disturbance of the adiabatic condition.

For both types of oscillations, the transverse focusing condition in the smooth approximation and the behavior of U_{eff} in the transverse-field undulator linac are similar to those in its longitudinal-field counterpart (see above). In the latter, fast oscillations of the velocities and particle coordinates have a significant effect on the transverse motion of the beam because of transverse particle–field interaction. The oscillations increase (three- or fourfold) the effective reduced transverse emittance of the beam in the acceleration mode, which decrease the current transmission coefficient when the channel aperture is limited.

NUMERICAL SIMULATION OF THE BEAM DYNAMICS IN THE UNDULATOR LINAC

To verify the analytical results for the beam dynamics in the rf undulator linac, as well as to find the limiting beam current, we performed a numerical simulation of the dynamics with allowance for the Coulomb field of the beam. The basic parameters of the accelerator and the simulation results are summarized in the table.

Consider first the results for the undulator linac with a longitudinal rf field ($\mu = \pi$). The numerical simulation of the deuterium ion dynamics in the smooth approximation shows that, at accelerator length $L = 2.5$ m and the effective amplitude of the combined wave $E_{\text{eff}} = e\lambda E_0 E_1/(2\pi W_0 \beta_{\text{in}}) = 30$ kV/cm, the beam output is 1.3–1.5 MeV and the current transmission coefficient $K_{\text{t}} = 90\%$. Particle losses are due to the disturbed adiabatic condition and can be depressed by tailoring the function of variation of the combined wave amplitude and phase $\psi_s(\xi)$ during bunching. Two bunches form during a period of the rf field (as also follows from Eq. (4) or (8)). When simulating the beam dynamics in the actual field, we found that fast velocity and coordinate oscillations substantially reduce current transmission coefficient K_{t}. It turned out that the maximal value of K_{t} is 60%. Furthermore, K_{t} considerably depends on amplitude ratio χ between the fundamental and first rf field harmonics. This fact confirms the analytical prediction of the presence of an optimal value of χ. In our simulation, optimal χ lies in the range 0.3–0.4. Such an amplitude ratio is quite feasible. The optimal length of the buncher is roughly equal to half the accelerator length.

In this undulator linac, the limiting cross-sectional size of the beam is 5×0.3 cm^2; the maximal current, $I_{\max} = 200$–250 mA (the maximal current density is $J_{\max} = 0.12$ A/cm^2). In the course of acceleration, the reduced effective emittance increases roughly twofold; however, the growth in this section is slower than in the bunching section. The cross-sectional size of the beam increases only slightly. A halo does not form. The maximal emittance of the beam calculated by numerical simulation agrees well with the channel acceptance found analytically.

The results of numerical simulation of the beam dynamics in the undulator linac with a transverse rf field for oscillations with $\mu = \pi$ differ insignificantly from the case considered above (longitudinal-field linac). The maximal cross-sectional size of the beam is somewhat larger, 7×0.3 cm^2; $K_{\text{t}} = 65\%$; optimal value of χ is 0.3–0.4; and the maximal beam current is also somewhat higher, 300–350 mA. It is worth noting that here particle losses due to longitudinal motion are absent during bunching.

For oscillations with $\mu = 0$, the current transmission coefficient at $\chi < 1$ is small in both types of the undulator linacs: for the longitudinal field, maximal $K_{\text{t}} = 30$–35%; for the transverse field, no higher than 10%. These data corroborate the statement that transverse focusing in the undulator linac with $\mu = 0$ is insufficient.

Basic parameters of the linac with an rf undulator

Parameters	Longitudinal undulator	Transverse undulator
Accelerator length L, m	2.5	
Injection energy of deuterium ions W_{in}, keV (β_{in})	100 (0.01)	
Amplitude E_0 of rf field fundamental harmonic, kV/cm	200	210
Amplitude E_1 of rf field first harmonic, kV/cm	80	70
Buncher-to-accelerator length ratio	0.5	0.5
Cross section of the undulator linac channel, cm	10×0.7	10×0.8
Initial beam size, cm^2	5×0.3	7×0.3
Transverse wavenumber ratio h_x/h_y	1/25	
Maximal transverse and longitudinal initial emittance		
ε_x, mm mrad	30π	30π
ε_y, mm mrad	0.7π	0.06π
ε_φ, keV mrad	25	40
Maximal transverse and longitudinal acceptance		
A_x, mm mrad	60π	60π
A_y, mm mrad	2π	2.5π
A_φ, keV mrad	40	40
Maximal beam current density $J_{\max}$, A/cm^2	0.12	
Maximal beam current $I_{\max}$, A	0.2–0.25	0.3–0.35
Current transmission coefficient K_{t}, %	60	65
Maximal beam energy $W_{\max}$, MeV ($\beta_{\max}$)	1.2–1.5 (0.034–0.04)	

The efficiency of such a linac can be raised by using tailored geometry and period of the accelerator channel so as to provide $\chi > 1$. For an accelerator length of 2.5 m, the maximal beam energy here is 0.9–1.1 MeV. Because of a low current transmission coefficient, the maximal current for the undulator linac with $\mu = 0$ was not estimated.

CONCLUSIONS

Thus, the analysis of the beam dynamics in an rf undulator linac shows that the new type of accelerator can be used to advantage in generating, bunching, and accelerating high-current ion beams. The undulator linac with field oscillations $\mu = \pi$ makes it possible to accelerate ribbon beams to energies of about 1.5 MeV with a high rate of acceleration. The maximal beam current may reach 350 mA. In this device, the rate of acceleration and the maximal current are higher than in the conventional RFQ accelerator. While the current transmission coefficient estimated in this study is no higher than 65%, it can be considerably increased by optimizing the function of variation of the equilibrium particle phase and field amplitude in the buncher with special methods.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, grant no. 04-01-16667.

REFERENCES

1. I. M. Kapchinskii and V. A. Teplyakov, Prib. Tekh. Éksp., No. 2, 19 (1970).
2. V. S. Tkalich, Zh. Éksp. Teor. Fiz. **32**, 625 (1957) [Sov. Phys. JETP **5**, 518 (1957)].
3. E. S. Masunov and N. E. Vinogradov, Zh. Tekh. Fiz. **71** (9), 79 (2001) [Tech. Phys. **46**, 1143 (2001)].
4. V. D. Danilov, A. A. Iliin, and Yu. K. Batygin, in *Proceedings of the 3rd European Particle Accelerator Conference (EPAC-92), Berlin, 1992*, Vol. 1, pp. 569–571.
5. E. S. Masunov, S. M. Polozov, and N. E. Vinogradov, Problemy Atomnoi Nauki i Tekhniki, No. 5, 71 (2001).
6. E. S. Masunov, Zh. Tekh. Fiz. **60** (8), 152 (1990) [Sov. Phys. Tech. Phys. **35**, 962 (1990)].
7. E. S. Masunov, Zh. Tekh. Fiz. **71** (11), 85 (2001) [Tech. Phys. **46**, 1433 (2001)].
8. E. S. Masunov and S. M. Polozov, in *Proceedings of the 9th International Workshop on Beam Dynamics and Optimization (BDO'2002), St.-Petersburg, 2002* (Izd. NIIKh, St. Petersburg, 2002), pp. 176–185.

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 937–943. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 119–125.
Original Russian Text Copyright © 2005 by Kolesnikov, Manuilov.

ELECTRON AND ION BEAMS, ACCELERATORS

Moment Equations and the Dynamic Equilibrium Condition for a Relativistic Electron Beam Propagating Along an External Magnetic Field in Dense and Rarefied Gas-Plasma Media

E. K. Kolesnikov and A. S. Manuilov
Smirnov Research Institute of Mathematics and Mechanics, St. Petersburg State University,
St. Petersburg, 198504 Russia
e-mail: Kolesnikov evg@mail.ru

Received October 29, 2004

Abstract—Equations of transfer of mass, momentum, and energy for a transverse segment of a paraxial relativistic electron beam propagating in dense and rarefied gas-plasma media along an external magnetic field are derived from the kinetic equation. A virial equation is obtained, and the dynamic equilibrium condition that generalizes the well-known Bennet equation for the cases under study is found. *© 2005 Pleiades Publishing, Inc.*

INTRODUCTION

Transport of relativistic electron beams (REBs) in dense and rarefied gas-plasma media has recently become a subject of extensive investigation [1–20]. Prominent among the problems concerned with REB propagation in the above media is the transverse dynamics of the beams. Earlier [12], the authors formulated a kinetic equation that describes the evolution of the distribution function for particles of a transverse segment of a paraxial monoenergetic axisymmetric REB propagating in dense and rarefied media along an external uniform magnetic field. In this work, we, based on this kinetic equation, derive equations of transfer of mass, momentum, and energy; a virial equation; and a dynamic equilibrium condition. The equations derived are a generalization of similar equations for an REB propagating in a dense gas plasma in the absence of a magnetic field [5] for the case when the beam is transported under ion focusing (IF) conditions in both a dense and rarefied gas plasma subjected to an axial uniform magnetic field.

PROBLEM DEFINITION

Consider an axisymmetric quasi-stationary REB with the symmetry axis directed along the z axis that is injected into a homogeneous gas plasma subjected to a permanent magnetic field aligned with the z axis (the magnetic induction is $\mathbf{B} = B_0\mathbf{i}_z$, where $\mathbf{i}_z$ is the unit vector along the z direction). We represent the beam as a set of thin transverse segments S^τ, each being injected at time instant $t = \tau$ (at $z = 0$) and containing a fixed number of particles.

We showed previously [12] that the beam particle longitudinal motion in any segment S^τ can be considered as determinate in the paraxial approximation. Unlike the longitudinal (axial) motion, the transverse dynamics of the particles is of stochastic character: in the phase state of transverse coordinates $\mathbf{r}_\perp$ and transverse momenta $\mathbf{p}_\perp$, the state of the beam can be characterized by distribution function $f^\tau(\mathbf{r}_\perp, \mathbf{p}_\perp, t)$. According to [5, 12], if collisions of beam particles with neutrals of a gas plasma result in multiple small-angle scattering and if the scattering is assumed to be elastic and isotropic, the evolution of distribution function $f^\tau(\mathbf{r}_\perp, \mathbf{p}_\perp, t)$ of particles entering into segment S^τ is described by the equation

$$\frac{\partial f^\tau}{\partial t} + \frac{\mathbf{p}_\perp}{\gamma m} \cdot \nabla_\perp f^\tau + [-e\nabla_\perp(\varphi_0 - \beta\mu A_z) + \Omega_b \mathbf{p}_\perp \times \mathbf{i}_z] \cdot \nabla_{\mathbf{p}_\perp} f^\tau = \frac{m\gamma S}{2}\Delta_{\mathbf{p}_\perp} f^\tau, \tag{1}$$

here $\Delta_\perp$ is the transverse component of the Laplacian.

In formula (1), coefficient μ is given by

$$\mu = 1 - \frac{1 - \alpha_c}{\beta^2(1 - \alpha_m)}, \tag{2}$$

where $\beta = v_z/c$ (v_z is the velocity axial component, which is assumed to be constant for all particles, since the beam is paraxial and monoenergetic, and c is the speed of light); α_c and α_m are, respectively, the degrees (coefficients) of charge and magnetic (current) neutralization of the beam; and φ_0 is a given potential of the ion background electric field (in the IF regime).

1063-7842/05/5007-0937$26.00 © 2005 Pleiades Publishing, Inc.

According to [12], when the beam is transported in a dense plasma, coefficients α_c and α_m may be assigned constant values and potential φ_0 may be set equal to zero, $\varphi_0 = 0$. In the IF regime, constants α_c and α_m, as well as φ_0, are zero. The other quantities in (1) have the following physical meaning: $\gamma = (1 - \beta^2)^{-1/2}$ is the Lorentz factor of beam particles; m and e are, respectively, the electron mass at rest and charge; $\Omega_b = |e|B_0/(\gamma mc)$ is the gyrofrequency of the beam in the external magnetic field; S is the mean rate of change of the kinetic energy of the particle transverse motion under multiple Coulomb scattering of the particles by atoms and molecules of the background gas (it is considered as a given characteristic of the scattering medium and beam [5, 12, 21–23]); and A_z is the z component of the potential of the collective electromagnetic field in the beam–plasma system. For the quasi-stationary paraxial beam being studied, A_x satisfies the equation

$$\Delta_\perp A_z = -\frac{4\pi}{c}(1-\alpha_m)J_{bz}, \tag{3}$$

where J_{bz} is the z component of the beam current density.

Let us introduce screening radius R_s for the self-consistent electromagnetic field of the background plasma, that is, assume that

$$\varphi_0\big|_{r \in R_s} = A_z\big|_{r \geq R_{R_s}} \equiv 0. \tag{4}$$

A solution to Eq. (3) satisfies boundary condition (4); that is,

$$A_z = -\frac{2}{c}I_b(1-\alpha_m) \times \int d\mathbf{r}'_\perp \ln\frac{|\mathbf{r}_\perp - \mathbf{r}'_\perp|}{R_s}\int d\mathbf{p}_\perp f^\tau(\mathbf{r}'_\perp, \mathbf{p}_\perp, t). \tag{5}$$

In view of (5), Eq. (1) can be viewed as an integro-differential equation for segment particle distribution function $f^\tau(\mathbf{r}_\perp, \mathbf{p}_\perp, t)$ that should be solved subject to

$$f^\tau(\mathbf{r}_\perp, \mathbf{p}_\perp, t)\big|_{t=\tau} = f_0(\mathbf{r}_\perp, \mathbf{p}_\perp, \tau), \tag{6}$$

where $f_0(\mathbf{r}_\perp, \mathbf{p}_\perp, \tau)$ is a given beam particle distribution over transverse coordinates.

MOMENT EQUATIONS

From Eq. (1), one can derive equations for the first moments of distribution function $f^\tau(\mathbf{r}_\perp, \mathbf{p}_\perp, t)$, which are the basic macroscopic characteristics of the beam: mass, momentum, and energy (the so-called transfer equations).

To derive moment equations, we suppose that nonzero values of distribution function f^τ occupy bounded domain Ω with boundary $\partial\Omega$ in the transverse momentum space; that is,

$$f^\tau\big|_{\mathbf{p}_\perp \in \Omega} \equiv 0. \tag{7}$$

Next, we take advantage of the Green's theorem

$$\int_\Gamma (v\Delta u - u\Delta v)d\mathbf{x} = \int_{\partial\Gamma}\left(v\frac{\partial u}{\partial \mathbf{n}} - u\frac{\partial v}{\partial \mathbf{n}}\right)\cdot dS^*, \tag{8}$$

where $\partial\Gamma$ is the boundary of domain $\Gamma \subset R^n$ and dS' is an elementary area of surface $\partial\Gamma$ that is assigned the direction of positive normal $\mathbf{n}$, and integral relationships

$$\int_\Gamma (v\nabla u + u\nabla v)d\mathbf{x} \equiv \int_{\partial\Gamma} uv\mathbf{n}dS^*, \tag{9}$$

$$\int_\Gamma (v\nabla\cdot\mathbf{w} + \mathbf{w}\cdot\nabla v)d\mathbf{x} = \int_{\partial\Gamma} v\mathbf{w}\cdot\mathbf{n}dS^*, \tag{10}$$

which are the corollaries of the gradient theorem and divergence theorem, respectively.

CONTINUITY EQUATION

Integrating kinetic equation (1) over the transverse momentum space yields

$$\begin{aligned}&\frac{\partial}{\partial t}\int f^\tau d\mathbf{p}_\perp + \nabla_\perp\cdot\int\frac{\mathbf{p}_\perp}{m\gamma}f^\tau d\mathbf{p}_\perp \\ &- e\nabla_\perp(\varphi_0 - \beta\mu A_z)\cdot\int\nabla_{\mathbf{p}_\perp}f^\tau d\mathbf{p}_\perp \\ &+ \Omega_b\int\left(p_y\frac{\partial f^\tau}{\partial p_x} - p_x\frac{\partial f^\tau}{\partial p_y}\right)d\mathbf{p}_\perp = \frac{m\gamma S}{2}\int\Delta_{\mathbf{p}_\perp}f^\tau d\mathbf{p}_\perp,\end{aligned} \tag{11}$$

where $\mathbf{p}_\perp = p_x\mathbf{i}_x + p_y\mathbf{i}_y$ $(\mathbf{i}_x \times \mathbf{i}_y = \mathbf{i}_z)$.

First of all, note that

$$\int f^\tau d\mathbf{p}_\perp = \chi(r_\perp, t), \quad \int\frac{\mathbf{p}_\perp}{m\gamma}f^\tau d\mathbf{p}_\perp = \chi\frac{\widetilde{\mathbf{p}_\perp}}{m\gamma}. \tag{12}$$

In view of (7) and (9), we have

$$\int\nabla_{\mathbf{p}_\perp}f^\tau d\mathbf{p}_\perp = \int\mathbf{n}f^\tau dS_{\mathbf{p}_\perp} \equiv 0, \tag{13}$$

and, by virtue of (7) and (10),

$$\begin{aligned}&\int\left(p_y\frac{\partial f^\tau}{\partial p_x} - p_x\frac{\partial f^\tau}{\partial p_y}\right)d\mathbf{p}_\perp \\ &= -\int f^\tau\nabla_{\mathbf{p}_\perp}\cdot\mathbf{w}_1(\mathbf{p}_\perp)d\mathbf{p}_\perp \equiv 0,\end{aligned} \tag{14}$$

where $\mathbf{w}_1 = p_y\mathbf{i}_x - p_x\mathbf{i}_y$ with $\nabla_{\mathbf{p}_\perp}\cdot\mathbf{w}_1 \equiv 0$.

Finally, with regard to (7) and (8), we arrive at the relationship

$$\int\Delta_{\mathbf{p}_\perp}f^\tau d\mathbf{p}_\perp = \int_{\partial\Omega}\frac{\partial f^\tau}{\partial\mathbf{n}}\cdot d\mathbf{S}^* \equiv 0. \tag{15}$$

Substituting (12)–(15) into (11), we obtain the continuity equation

$$\frac{\partial\chi}{\partial t} + \nabla_\perp \cdot \left(\frac{\chi\widetilde{\mathbf{p}_\perp}}{m\gamma}\right) = 0, \tag{16}$$

where $\chi(\mathbf{r}_\perp, t)$ is the beam particle density in segment S^τ, which is given by the integral in (12), and

$$\widetilde{\mathbf{p}_\perp}(\mathbf{r}_\perp, t) = \frac{1}{\chi}\int \mathbf{p}_\perp f^\tau d\mathbf{p}_\perp \tag{17}$$

is the mean transverse momentum of the particles in segment S^τ.

Note that, since $\tilde{\mathbf{p}}_\perp/(m\gamma) = \tilde{\mathbf{v}}_\perp$ in (16), where $\tilde{\mathbf{v}}_\perp$ is the mean transverse velocity of REB particles, Eq. (16) is, in essence, an ordinary continuity equation that describes the particle conservation law in the segment considered.

EQUATION OF MOMENTUM TRANSFER

Multiplying Eq. (1) by $\mathbf{p}_x$ and integrating the result over transverse momenta $\mathbf{p}_\perp$ yields

$$\begin{gathered}\frac{\partial}{\partial t}\int p_x f^\tau d\mathbf{p}_\perp + \nabla_\perp \cdot \int \frac{p_x\mathbf{p}_\perp}{m\gamma} f^\tau d\mathbf{p}_\perp \\ - e\nabla_\perp(\varphi_0 - \beta\mu A_z) \cdot \int p_x \nabla_{\mathbf{p}_\perp} f^\tau d\mathbf{p}_\perp \\ + \Omega_\mathrm{b}\int p_x\left(p_y\frac{\partial f^\tau}{\partial p_x} - p_x\frac{\partial f^\tau}{\partial p_y}\right)d\mathbf{p}_\perp = \frac{\gamma m S}{2}\int p_x \Delta_{\mathbf{p}_\perp} f^\tau d\mathbf{p}_\perp.\end{gathered} \tag{18}$$

Consider the integrals on the left and right of Eq. (18),

$$p_x f^\tau d\mathbf{p}_\perp = \chi\widetilde{p_x}; \quad \int \frac{p_x\mathbf{p}_\perp}{m\gamma} f^\tau d\mathbf{p}_\perp = \chi\frac{\widetilde{p_x\mathbf{p}_\perp}}{m\gamma}. \tag{19}$$

In view of (7) and (9), we have

$$\begin{aligned}\int p_x \nabla_{\mathbf{p}_\perp} f^\tau d\mathbf{p}_\perp &= -\int f^\tau \nabla_{\mathbf{p}_\perp} p_x d\mathbf{p}_\perp \\ &= -\int_\Omega f^\tau \mathbf{i}_x d\mathbf{p}_\perp = -\mathbf{i}_x\chi,\end{aligned} \tag{20}$$

and, in view of (7) and (10),

$$\begin{aligned}\int p_x\left(p_y\frac{\partial f^\tau}{\partial p_x} - p_x\frac{\partial f^\tau}{\partial p_y}\right)d\mathbf{p}_\perp &= \int \mathbf{w}_2(\mathbf{p}_\perp) \cdot \nabla_{\mathbf{p}_\perp} f^\tau d\mathbf{p}_\perp \\ = -\int f^\tau \nabla_{\mathbf{p}_\perp} \cdot \mathbf{w}_2 d\mathbf{p}_\perp &= -\int p_y f^\tau d\mathbf{p}_\perp = -\chi\widetilde{p_y},\end{aligned} \tag{21}$$

where $\mathbf{w}_2(\mathbf{p}_\perp) = p_x p_y \mathbf{i}_x - p_x^2 t_y$ with $\nabla_{\mathbf{p}_\perp} \cdot \mathbf{w}_2 = p_y$.

Finally, with regard to (7) and (8), we get

$$\int p_x \Delta_{\mathbf{p}_\perp} f^\tau d\mathbf{p}_\perp = \int_{\partial\Omega}\left(p_x\frac{\partial f^\tau}{\partial \mathbf{n}} - f^\tau\frac{\partial p_x}{\partial \mathbf{n}}\right) \cdot d\mathbf{S}_{\mathbf{p}_\perp} \equiv 0. \tag{22}$$

Substituting (19)–(22) into Eq. (18) yields

$$\begin{gathered}\frac{\partial\chi\widetilde{p_x}}{\partial t} + \nabla_\perp \cdot \left(\chi\frac{\widetilde{p_x\mathbf{p}_\perp}}{m\gamma}\right) \\ + e\chi\frac{\partial}{\partial x}(\varphi_0 - \beta\mu A_z) - \chi\Omega_\mathrm{b}\widetilde{p_y} = 0.\end{gathered} \tag{23}$$

Then, multiplying Eq. (1) by p_y, integrating the result over transverse momenta $\mathbf{p}_\perp$, and performing similar rearrangements, we obtain

$$\begin{gathered}\frac{\partial\chi\widetilde{p_y}}{\partial t} + \nabla_\perp \cdot \left(\chi\frac{\widetilde{p_y\mathbf{p}_\perp}}{m\gamma}\right) \\ + e\chi\frac{\partial}{\partial y}(\varphi_0 - \beta\mu A_z) - \chi\Omega_\mathrm{b}\widetilde{p_x} = 0.\end{gathered} \tag{24}$$

Equations (23) and (24) can be written in the vector form,

$$\begin{gathered}\frac{\partial\chi\widetilde{\mathbf{p}_\perp}}{\partial t} + \nabla_\perp \cdot \left(\chi\frac{\widetilde{\mathbf{p}_\perp\mathbf{p}_\perp}}{m\gamma}\right) \\ + e\chi\nabla_\perp(\varphi_0 - \beta\mu A_z) + \chi\Omega_\mathrm{b}(\mathbf{i}_z \times \tilde{\mathbf{p}}_\perp) = 0.\end{gathered} \tag{25}$$

Let us transform Eq. (25) into the form allowing a simple physical interpretation. Transverse momentum $\mathbf{p}_\perp$ is represented as

$$\mathbf{p}_\perp = \widetilde{\mathbf{p}_\perp} + \delta\mathbf{p}_\perp. \tag{26}$$

Then,

$$\begin{gathered}\nabla_\perp \cdot (\chi\widetilde{\mathbf{p}_\perp\mathbf{p}_\perp}) = \nabla_\perp \cdot (\chi\mathbf{p}_\perp\mathbf{p}_\perp) \\ + \nabla_\perp \cdot (\chi\widetilde{\delta\mathbf{p}_\perp\delta\mathbf{p}_\perp}) + 2\nabla_\perp \cdot (\chi\mathbf{p}_\perp\delta\mathbf{p}_\perp).\end{gathered} \tag{27}$$

From (26), it follows that $\widetilde{\delta\mathbf{p}_\perp} = 0$. In addition, we can write

$$\chi\widetilde{\delta\mathbf{p}_\perp\delta\mathbf{p}_\perp} = m\gamma\int \delta\mathbf{p}_\perp\delta\mathbf{v}_\perp f^\tau d\mathbf{p}_\perp = m\gamma\tilde{\tilde{P}}_\perp, \tag{28}$$

where

$$\tilde{\tilde{P}}_\perp = \int(\mathbf{p}_\perp - \widetilde{\mathbf{p}_\perp})(\mathbf{v}_\perp - \widetilde{\mathbf{v}_\perp}) f^\tau d\mathbf{p}_\perp \tag{29}$$

is the stress tensor.

Finally, the first term on the right of (27) is written in the form

$$\nabla_\perp \cdot (\chi\widetilde{\mathbf{p}_\perp\mathbf{p}_\perp}) = \widetilde{\mathbf{p}_\perp}\nabla_\perp \cdot (\chi\widetilde{\mathbf{p}_\perp}) + \chi(\widetilde{\mathbf{p}_\perp} \cdot \nabla_\perp)\widetilde{\mathbf{p}_\perp}. \tag{30}$$

Them, with regard to (27)–(30); continuity equation (16); and the relationship $\widetilde{\mathbf{p}_\perp} = m\gamma\widetilde{\mathbf{v}_\perp}$, which follows from (25), we arrive at the equation

$$\left(\frac{\partial}{\partial t}+\widetilde{\mathbf{v}_\perp}\cdot\nabla_\perp\right)\widetilde{\mathbf{p}_\perp} = -\frac{\nabla_\perp\cdot\tilde{\tilde{P}}_\perp}{\chi}+e\left[\mathbf{E}_\perp^{\text{eff}}+\frac{1}{c}(\widetilde{\mathbf{v}_\perp}\times\mathbf{B}_0)\right], \tag{31}$$

where $\mathbf{E}_\perp^{\text{eff}} = -\nabla_\perp(\varphi_0-\beta\mu A_z)$ is the transverse component of the effective electric field.

The terms of (31) can be interpreted as follows. The left of (31), $(\partial/\partial t + \mathbf{v}_\perp\cdot\nabla_\perp)\tilde{\mathbf{p}}_\perp = d\tilde{\mathbf{p}}_\perp/dt$, is the total derivative of momentum $\tilde{\mathbf{p}}_\perp$ along the path of a particle moving with velocity $\tilde{\mathbf{v}}_\perp$. The first term on the right of (31) is the hydrodynamic force due to a spread in the transverse momenta of beam particles; the second one, the electromagnetic force due to effective electric field $\mathbf{E}_\perp^{\text{eff}}$ and external magnetic field.

EQUATION OF ENERGY TRANSFER

Multiplying kinetic equation (1) by $p_\perp^2/(2m\gamma)$ and integrating the result over the transverse momentum space yields (with regard to the explicit time dependence of Lorentz factor γ)

$$\frac{\partial}{\partial t}\int\frac{p_\perp^2}{2m\gamma}f^\tau d\mathbf{p}_\perp+\frac{1}{\gamma}\frac{d\gamma}{dt}\int\frac{p_\perp^2}{2m\gamma}f^\tau\mathbf{p}_\perp+\nabla_\perp\cdot\int\frac{\mathbf{p}_\perp p_\perp^2}{2m^2\gamma^2}f^\tau d\mathbf{p}_\perp$$
$$-e\nabla_\perp(\varphi_0-\beta\mu A_z)\cdot\int\frac{p_\perp^2}{2m\gamma}\nabla_{\mathbf{p}_\perp}f^\tau d\mathbf{p}_\perp \tag{32}$$
$$+\frac{\Omega_{\text{b}}}{2m\gamma}\int p_\perp^2\left(p_y\frac{\partial f^\tau}{\partial p_x}-p_x\frac{\partial f^\tau}{\partial p_y}\right)d\mathbf{p}_\perp = \frac{S}{4}\int p_\perp^2\Delta_{\mathbf{p}_\perp}f^\tau d\mathbf{p}_\perp.$$

For the integrals entering into Eq. (32), we have

$$\int\frac{p_\perp^2}{2m\gamma}f^\tau d\mathbf{p}_\perp = \chi\frac{\widetilde{p_\perp^2}}{2m\gamma}, \tag{33}$$

$$\int\frac{\mathbf{p}_\perp p_\perp^2}{2m^2\gamma^2}f^\tau dp_\perp = \chi\frac{\widetilde{\mathbf{p}_\perp p_\perp^2}}{2m^2\gamma^2} = \chi\frac{\widetilde{\mathbf{v}_\perp p_\perp^2}}{2m\gamma}. \tag{34}$$

In view of (7), (9), and relationship $\nabla_{\mathbf{p}_\perp} = 2\mathbf{p}_\perp$, we come to

$$\int\frac{p_\perp^2}{2m\gamma}\nabla_{p_\perp}f^\tau d\mathbf{p}_\perp = -\frac{1}{m\gamma}\int\mathbf{p}_\perp f^\tau dp_\perp = -\chi\frac{\widetilde{\mathbf{p}_\perp}}{m\gamma}. \tag{35}$$

Taking into account (7) and (10) gives

$$\int p_\perp^2\left(p_y\frac{\partial f^\tau}{\partial p_x}-p_x\frac{\partial f^\tau}{\partial p_y}\right)d\mathbf{p}_\perp = \int\mathbf{w}_3(\mathbf{p}_\perp)\cdot\nabla_{\mathbf{p}_\perp}f^\tau d\mathbf{p}_\perp$$
$$= -\int f^\tau\nabla_{p_\perp}\cdot\mathbf{w}_3(\mathbf{p}_\perp)d\mathbf{p}_\perp = 0, \tag{36}$$

where $\mathbf{w}_3(\mathbf{p}_\perp) = p_\perp^2p_y\mathbf{i}_x - p_\perp^2p_x\mathbf{i}_y$ with $\nabla_{\mathbf{p}_\perp}\cdot\mathbf{w}_3(\mathbf{p}_\perp)\equiv 0$.

Finally, with (7), (8), and relationship $\Delta_{p_\perp}p_\perp^2 = 4$, we get

$$\int p_\perp^2\Delta_{\mathbf{p}_\perp}f^\tau d\mathbf{p}_\perp = 4\int f^\tau d\mathbf{p}_\perp = 4\chi. \tag{37}$$

Substituting (33)–(37) into (32) yields the equation of energy transfer

$$\frac{\partial}{\partial t}\left(\frac{\chi\widetilde{p_\perp^2}}{2m\gamma}\right) = -\nabla_\perp\cdot\left(\chi\frac{\widetilde{\mathbf{p}_\perp p_\perp^2}}{2m^2\gamma^2}\right)$$
$$-\frac{e\chi\widetilde{\mathbf{p}_\perp}}{m\gamma}\cdot\nabla_\perp(\varphi_0-\beta\mu A_z)-\frac{1}{\gamma}\frac{d\gamma}{dt}\frac{\chi\widetilde{p_\perp^2}}{2m\gamma}+\chi S. \tag{38}$$

The first term on the right of Eq. (38) is the rate of variation of the mean energy of the particles from elementary segment S^τ that move in the transverse direction. The variation is associated with the energy flux with density

$$q = \frac{\chi\widetilde{\mathbf{p}_\perp p_\perp^2}}{2m^2\gamma^2} = \frac{\chi\widetilde{\mathbf{v}_\perp p_\perp^2}}{2m\gamma}. \tag{39}$$

The second term on the right of (38) can be written in the form

$$-\frac{e\chi\widetilde{\mathbf{p}_\perp}}{m\gamma}\cdot\nabla_\perp(\varphi_0-\beta\mu A_z) = \mathbf{j}_\perp\cdot\mathbf{E}_\perp^{\text{eff}}, \tag{40}$$

where $\mathbf{j}_\perp = e\chi\widetilde{\mathbf{p}_\perp}/(m\gamma) = e\chi\widetilde{\mathbf{v}_\perp}$.

As follows from (40), this term characterizes the rate of variation of the transverse motion energy that is associated with the work of self-consistent electromagnetic field forces acting on the particles.

Finally, the two last terms on the right of (38) characterize the rate of variation of the transverse motion energy due to, respectively, inelastic and elastic collisions between particles of the beam and environment.

To conclude, we note that the above procedure can be easily extended to the case when the domain of nonzero distribution function f^τ in the transverse momentum space is unbounded but f^τ fairly rapidly tends to zero as $p_\perp \longrightarrow 0$. Here, integration in (11), (18), and (34) is first over bounded domain Ω (using integral relationships (8)–(10)) and then over the transverse momentum space with boundary ∂Ω of domain Ω tending to infinity. It is easy to check that such a procedure yields moment equations (16), (25), and (38) iff distribution function f^τ decays with increasing $p_\perp$ faster than $1/p_\perp^4$. In this case, all the integrals over ∂Ω at $\partial\Omega \longrightarrow \infty$ vanish and all the integrals over the transverse momentum space converge.

VIRIAL EQUATION, INTEGRAL OF MEAN GENERALIZED ANGULAR MOMENTUM, AND DYNAMIC EQUILIBRIUM CONDITION

Scalarly multiplying the equation of momentum transfer (Eq. (25)) by $-r_\perp/2$ and integrating the resulting expression over transverse coordinates yields

$$\begin{aligned} &-\frac{1}{2}\frac{d}{dt}\int \chi \mathbf{p}_\perp \cdot \mathbf{r}_\perp d\mathbf{r}_\perp \\ &-\frac{1}{2}\int \nabla_\perp \left(\frac{\chi \widetilde{\mathbf{p}_\perp \mathbf{p}_\perp}}{m\gamma}\right) \cdot \mathbf{r}_\perp d\mathbf{r}_\perp = V, \end{aligned} \tag{41}$$

where

$$V \equiv -\frac{1}{2}\chi \mathbf{r}_\perp \cdot [-e\nabla_\perp(\varphi_0 - \beta\mu A_z) + \Omega_\mathrm{b}(\widetilde{\mathbf{p}}_\perp \times \mathbf{i}_z)]d\mathbf{r}_\perp \tag{42}$$

is the mean virial.

When finding the integrals in (41) and (42), we assume that the particles from the beam segment considered remain within bounded domain Γ with boundary $\partial\Gamma$ in the space of transverse coordinates $\mathbf{r}_\perp$ at any time; hence, function χ satisfies the condition

$$\chi(\mathbf{r}_\perp, t)\big|_{\mathbf{r}_\perp \in \Gamma} \equiv 0. \tag{43}$$

Consider first the integral

$$\begin{aligned} \int \chi \widetilde{\mathbf{p}}_\perp \cdot \mathbf{r}_\perp d\mathbf{r}_\perp &= \frac{1}{2}\int \nabla_\perp \cdot (\chi \widetilde{\mathbf{p}}_\perp r^2) d\mathbf{r}_\perp \\ &-\frac{1}{2}\int r^2 \nabla_\perp \cdot (\chi \widetilde{\mathbf{p}}_\perp) d\mathbf{r}_\perp. \end{aligned} \tag{44}$$

By virtue of the divergence theorem and condition (43), the first integral on the right of (44) can be written as

$$\int \nabla_\perp \cdot (\chi \widetilde{\mathbf{p}}_\perp r^2) d\mathbf{r}_\perp = \int_{\partial\Gamma} \chi r^2 \widetilde{\mathbf{p}}_\perp \cdot \mathbf{n} dS_{r_\perp} \equiv 0, \tag{45}$$

and the second integral (in view of continuity equation (16)) as

$$-\frac{1}{2}\int r^2 \nabla_\perp \cdot (\chi \tilde{\mathbf{p}}_\perp) d\mathbf{r}_\perp = \frac{m\gamma}{4}\frac{d\Re^2}{dt}, \tag{46}$$

where

$$\Re^2 \equiv 2\int \chi r^2 d\mathbf{r}_\perp. \tag{47}$$

The second integral on the left of (41) can be transformed as follows:

$$\begin{aligned} &\int \nabla_\perp \left(\frac{\chi \widetilde{\mathbf{p}_\perp \mathbf{p}_\perp}}{m\gamma}\right) \cdot \mathbf{r}_\perp d\mathbf{r}_\perp \\ &= \int_{\partial\Gamma} \frac{x\chi \widetilde{p_x \mathbf{p}_\perp}}{m\gamma} \cdot \mathbf{n} dS_{r_\perp} + \int_{\partial\Gamma} \frac{y\chi \widetilde{p_y \mathbf{p}_\perp}}{m\gamma} \cdot \mathbf{n} dS_{r_\perp} \qquad (48) \\ &-\int \frac{\chi(\widetilde{p_x^2} + \widetilde{p_y^2})}{m\gamma} d\mathbf{r}_\perp = -\int \frac{\chi \widetilde{p_\perp^2}}{m\gamma} d\mathbf{r}_\perp. \end{aligned}$$

Substituting (48) and (44) into (41) with regard to (45) and (46), we get

$$E_\perp - \frac{d}{dt}\left(\frac{m\gamma}{8}\frac{d\Re^2}{dt}\right) = V, \tag{49}$$

where $E_\perp \equiv \int \chi p_\perp^2/(2m\gamma) dr_\perp$ is the mean kinetic energy of the transverse motion of the particles from segment S^τ and mean virial V is defined by (42).

Consider now the integrals appearing in virial expression (42). Note first of all that

$$-\frac{1}{2}e\beta\mu \int \chi \mathbf{r}_\perp \cdot \nabla_\perp A_z d\mathbf{r}_\perp = -e\beta\mu\pi \int_0^\infty \chi r^2 \frac{dA_z}{dr} dr, \tag{50}$$

$$\frac{1}{2}e\int \chi \mathbf{r}_\perp \cdot \nabla_\perp \varphi_0 = e\pi \int_0^\infty \chi r^2 \frac{d\varphi_0}{dr} \tag{51}$$

by virtue of the axial symmetry of the problem.

To determine derivatives dA_z/dr and $d\varphi_0/dr$ in (50) and (51), we consider the Poisson equations for potentials A_z and φ_0,

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dA_z}{dr}\right) = -4\pi e\beta(1 - \alpha_\mathrm{m})\hat{N}_\mathrm{b}\chi(r), \tag{52}$$

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{d\varphi_0}{dr}\right) = 4\pi e\hat{N}_\Phi \chi_\Phi(r), \tag{53}$$

where $\hat{N}_\mathrm{b}$ and $\hat{N}_\Phi$ are the total linear densities of particles in the beam and ions in the ion channel, respectively, and $\chi_\Phi = n_\Phi(r)/\hat{N}_\Phi$ is a given function characterizing the ion radial distribution ($\int \chi_\Phi(r) dr_\perp = 1$).

Integrating (52) and (53) subject to $rdA_z/dr|_{r=0} = rd\varphi_{0z}/dr|_{r=0} = 0$ yields

$$\frac{dA_z}{dr} = -2e\beta(1 - \alpha_\mathrm{m})\frac{N_\mathrm{b}(r)}{r}, \tag{54}$$

$$\frac{d\varphi_0}{dr} = 2e\frac{N_\Phi(r)}{r}, \tag{55}$$

where

$$N_\mathrm{b}(r) = 2\pi\hat{N}_\mathrm{b}\int_0^r r'\chi(r')dr', \tag{56}$$

$$N_\Phi(r) = 2\pi\hat{N}_\Phi\int_0^r r'\chi_\Phi(r')dr'. \tag{57}$$

As follows from (56) and (57), $N_\mathrm{b}(r)$ and $N_\Phi(r)$ represent the linear densities of particles in the beam and ions in the ionic background in a tube of radius r.

Substituting (54) and (56) into the right of (50) gives

$$-\frac{1}{2}e\beta\mu\int\chi\mathbf{r}_\perp\cdot\nabla_\perp A_z d\mathbf{r}_\perp = e^2\beta^2\kappa\hat{N}_\mathrm{b}4\pi^2\int_0^\infty dr r\chi(r)\int_0^r dr'r'\chi(r'), \tag{58}$$

where constant $\kappa = (1-\alpha_\mathrm{m})\mu = (1-\alpha_\mathrm{m}) - (1-\alpha_\mathrm{c})/\beta^2$ (parameter μ is defined by (2)).

Finally, we note that the equality

$$4\pi^2\int_0^\infty dr r\chi(r)\int_0^r dr'r'\chi(r') = 2\pi^2\int_0^\infty dr r\chi(r)\int_0^\infty dr'r'\chi(r') = \frac{1}{2}\int d\mathbf{r}_\perp\chi(r)\int\mathbf{r}'_\perp\chi(r') = \frac{1}{2} \tag{59}$$

is valid with regard to (56).

Using (59), we eventually obtain

$$-\frac{1}{2}e\beta\mu\int\chi\mathbf{r}_\perp\cdot\nabla_\perp A_z d\mathbf{r}_\perp = \frac{e^2\beta^2\kappa\hat{N}_\mathrm{b}}{2}. \tag{60}$$

Similarly, substituting (55) and (57) into the right of (51) yields

$$\frac{e}{2}\int\chi\mathbf{r}_\perp\cdot\nabla_\perp\varphi_0 d\mathbf{r}_\perp = e^2 2\pi\int_0^\infty r\chi(r)N_\Phi(r)dr = e^2\int\chi(r)N_\Phi(r)d\mathbf{r}_\perp = e^2\langle N_\Phi(r)\rangle. \tag{61}$$

A number of practically interesting specific cases where $\langle N_\Phi(r)\rangle$ takes a very simple form deserve consideration here. If, in the domain occupied by the beam, the volume density of ions is constant along the radius and equals n_Φ^0, one can write

$$\langle N_\Phi(r)\rangle = \frac{\pi}{2}n_\Phi^0\Re^2, \tag{62}$$

where $\Re^2$ is given by integral (47).

If the ions have a δ-shaped distribution with linear density $\hat{N}_\Phi$ along the beam axis, we have

$$\langle N_\Phi(r)\rangle = \hat{N}_\Phi. \tag{63}$$

Finally, if the radial profiles of the beam and ionic background are the same, $\chi(r) = \chi_\Phi(r)$, we get

$$\langle N_\Phi(r)\rangle = \frac{\hat{N}_\Phi}{2}. \tag{64}$$

The second integral in expression (42) for the mean virial can be written in the form

$$-\frac{\Omega_\mathrm{b}}{2}\int\chi\mathbf{r}_\perp\cdot(\widetilde{\mathbf{p}_\perp}\times\mathbf{i}_z)d\mathbf{r}_\perp = -\frac{\Omega_\mathrm{b}}{2}\int\chi(\mathbf{r}_\perp\times\widetilde{\mathbf{p}_\perp})\cdot\mathbf{i}_z d\mathbf{r}_\perp = -\frac{\Omega_\mathrm{b}}{2}L, \tag{65}$$

where

$$L = \int\chi(\mathbf{r}_\perp\times\widetilde{\mathbf{p}_\perp})\cdot\mathbf{i}_z d\mathbf{r}_\perp = \int\chi r\tilde{p}_\theta d\mathbf{r}_\perp \equiv \langle rp_\theta\rangle \tag{66}$$

is the mean angular momentum of a particle from the segment considered.

Substituting (60), (61), and (65) into (42) yields an expression for mean virial V,

$$V = e^2\left(\frac{\beta^2\kappa\hat{N}_\mathrm{b}}{2} + \langle N_\Phi(r)\rangle\right) - \frac{\Omega_\mathrm{b}L}{2}, \tag{67}$$

where $\hat{N}_\mathrm{b}$ is the total linear density of beam particles, $\kappa = (1-\alpha_\mathrm{m}) - (1-\alpha_\mathrm{c})/\beta^2$, and $\langle N_\Phi(r)\rangle$ is defined by (61).

Now consider the derivative

$$\frac{dL}{dt} = \int r\frac{\partial(\chi\widetilde{p_\theta})}{\partial t}d\mathbf{r}_\perp. \tag{68}$$

Taking into account the equation of momentum transfer (see (25)) and the axial symmetry of the beam, we have

$$\int r\frac{\partial(\chi\widetilde{p_\theta})}{\partial t}d\mathbf{r}_\perp = \Omega_\mathrm{b}\int\chi r(\widetilde{\mathbf{p}_\perp}\times\mathbf{i}_z)\cdot\mathbf{i}_\theta d\mathbf{r}_\perp = -\Omega_\mathrm{b}\int\chi r\widetilde{p_r}d\mathbf{r}_\perp = -\Omega_\mathrm{b}\int\mathbf{r}_\perp\cdot(\chi\widetilde{\mathbf{p}_\perp})d\mathbf{r}_\perp. \tag{69}$$

Using continuity condition (16) and boundary condition (43), we get

$$\int\mathbf{r}_\perp\cdot(\chi\widetilde{\mathbf{p}_\perp})d\mathbf{r}_\perp = \frac{m\gamma}{2}\int\nabla_\perp r^2\cdot\left(\frac{\chi\widetilde{\mathbf{p}_\perp}}{m\gamma}\right)d\mathbf{r}_\perp = -\frac{m\gamma}{2}\int r^2\nabla_\perp\cdot\left(\frac{\chi\widetilde{\mathbf{p}_\perp}}{m\gamma}\right)d\mathbf{r}_\perp = \frac{m\gamma}{4}\frac{d\Re^2}{dt}. \tag{70}$$

In view of (70), Eq. (68) takes the form

$$\frac{dL}{dt} = -\Omega_\mathrm{b}\frac{m\gamma}{4}\frac{d\Re^2}{dt}; \tag{71}$$

whence, it follows that

$$\frac{d}{dt}\left(L + \Omega_\mathrm{b}\frac{m\gamma}{4}\Re^2\right) = 0. \tag{72}$$

It is easy to check that the integral of (72) is nothing else than the law of conservation of generalized angular momentum of a particle from segment S^τ. Indeed, in the paraxial approximation used in this work, generalized angular momentum is $\tilde{p}_\theta = rp_\theta + erA_\theta/c = rp_\theta +$

$\Omega_b m\gamma r^2/2$ and its mean is

$$\begin{aligned}\tilde{P}_\theta &= \int f^\tau\left(rp_\theta + \frac{\Omega_b m\gamma r^2}{2}\right)d\mathbf{r}_\perp d\mathbf{p}_\perp \\ &= \int\chi r\tilde{p}_\theta d\mathbf{r}_\perp + \frac{\Omega_b m\gamma}{2}\int\chi r^2 d\mathbf{r}_\perp = L + \frac{\Omega_b m\gamma}{4}\Re^2,\end{aligned} \tag{73}$$

which completely coincides with the parenthesized expression in (72).

The integral of (72) relates a current value of the mean of angular momentum L and quantity $\Re^2$,

$$L = L_0 + \frac{m\gamma\Omega_b}{4}(\Re_0^3 - \Re^2), \tag{74}$$

where L_0 and $\Re_0$ are the initial values of the corresponding quantities.

Substituting expression (67) into virial equation (51) yields

$$\begin{aligned}&\frac{d}{dt}\left(\frac{m\gamma}{8}\frac{d\Re^2}{dt}\right) \\ &= E_\perp - e^2\left(\frac{k\beta^2\hat{N}_b}{2} + \langle N_\Phi(\mathbf{r}_\perp)\rangle\right) + \frac{\Omega_b L}{2}.\end{aligned} \tag{75}$$

Assuming that $d\Re^2/dt \equiv 0$ in (75), we find the necessary condition for dynamic equilibrium of the beam segment studied,

$$E_\perp = \kappa T_B + e^2\langle N_\Phi(\mathbf{r}_\perp)\rangle - \frac{\Omega_b L}{2}, \tag{76}$$

where $T_B = e^2\beta^2 N_b/2 = e\beta J_{bz}/2c$ is the so-called Bennet temperature.

Condition (76) is an extension of the well-known Bennet equilibrium condition [5] for the case when an external magnetic field and neutralizing ionic background are present.

REFERENCES

1. A. A. Rukhadze, L. S. Bogdankevich, S. E. Rosinskii, and V. G. Rukhlin, *The Physics of High-Current Relativistic Electron Beams* (Atomizdat, Moscow, 1980) [in Russian].
2. A. N. Didenko, V. P. Grigor'ev, and Yu. P. Usov, *High-Power Electronic Beams and Their Applications* (Atomizdat, Moscow, 1977) [in Russian].
3. M. V. Kuzelev and A. A. Rukhadze, *Electrodynamics of Dense Electron Beams in Plasmas* (Nauka, Moscow, 1990) [in Russian].
4. R. B. Miller, *Introduction to the Physics of Intense Charged Particle Beams* (Plenum, New York, 1982; Mir, Moscow, 1984).
5. E. P. Lee, Phys. Fluids **19**, 60 (1976).
6. E. P. Lee, Phys. Fluids **21**, 1327 (1978).
7. H. L. Buchanan, Phys. Fluids **30**, 221 (1987).
8. E. R. Nadezhdin and G. A. Sorokin, Fiz. Plazmy **9**, 988 (1983) [Sov. J. Plasma Phys. **9**, 576 (1983)].
9. L. V. Glazychev and G. A. Sorokin, Fiz. Plazmy **16**, 592 (1990) [Sov. J. Plasma Phys. **16**, 340 (1990)].
10. E. R. Nadezhdin, Fiz. Plazmy **17**, 327 (1991) [Sov. J. Plasma Phys. **17**, 194 (1991)].
11. R. F. Fernsler, R. F. Hubbard, and M. Lampe, J. Appl. Phys. **75**, 3278 (1994).
12. E. K. Kolesnikov and A. S. Manuilov, Zh. Tekh. Fiz. **74** (9), 103 (2004) [Tech. Phys. **49**, 1208 (2004)].
13. E. K. Kolesnikov and A. D. Savkin, Pis'ma Zh. Tekh. Fiz. **20** (1), 54 (1994) [Tech. Phys. Lett. **20**, 26 (1994)].
14. E. K. Kolesnikov and A. S. Manuilov, Radiotekh. Élektron. (Moscow) **37**, 694 (1992).
15. E. K. Kolesnikov and A. S. Manuilov, Zh. Tekh. Fiz. **70** (5), 68 (2000) [Tech. Phys. **45**, 591 (2000)].
16. E. K. Kolesnikov and A. S. Manuilov, Zh. Tekh. Fiz. **70** (7), 127 (2000) [Tech. Phys. **45**, 939 (2000)].
17. A. S. Manuilov, Zh. Tekh. Fiz. **70** (1), 76 (2000) [Tech. Phys. **45**, 74 (2000)].
18. E. R. Nadezhdin and G. A. Sorokin, Fiz. Plazmy **14**, 619 (1988) [Sov. J. Plasma Phys. **14**, 365 (1988)].
19. E. K. Kolesnikov, Fiz. Plazmy **28**, 360 (2002) [Plasma Phys. Rep. **28**, 327 (2002)].
20. E. K. Kolesnikov and A. S. Manuilov, Zh. Tekh. Fiz. **67** (11), 62 (1997) [Tech. Phys. **42**, 1297 (1997)].
21. R. L. Liboff, *Introduction to the Theory of Kinetic Equations* (Wiley, New York, 1969; Mir, Moscow, 1974).
22. S. Chandrashekhar, *Stochastic Problems in Physics and Astronomy* (Rev. Mod. Phys. **15** (1943); GIIL, Moscow, 1947).
23. C. Cercignani, *Theory and Application of the Boltzmann Equation* (Springer, New York, 1988; Mir, Moscow, 1978).

Translated by V. Isaakyan

Technical Physics, Vol. 50, No. 7, 2005, pp. 944–947. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 126–130.
Original Russian Text Copyright © 2005 by Bocharov, Eletskii.

SURFACE, ELECTRON AND ION EMISSION

Effect of Screening on the Emissivity of Field Electron Emitters Based on Carbon Nanotubes

G. S. Bocharov* and A. V. Eletskii**

** Moscow Power Engineering Institute (Technical University),*
ul. Krasnokazarmennaya 14, Moscow, 111250 Russia
e-mail: www.mpei.ru
*** Russian Research Center Kurchatov Institute,*
pl. Akademika Kurchatova 1, Moscow, 123182 Russia
e-mail: www.kiae.ru

Received September 21, 2004

Abstract—The effect of screening on the emissivity of a field cathode built around a carbon nanotube array is analyzed. A numerical method of solving the Laplace equation for intricate-shape cathodes is developed that makes it possible to relate the amplification factor to the nanotube spacing in arrays containing as many as 225 emitters. Mutual screening of the tubes, which shows up in the dependence of the field amplification factor on the average emitter spacing, is studied numerically. The optimal spacing between the tubes that provides an emission current maximum density at a given applied voltage is determined. The role of edge effects in carbon nanotube screening is established. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

Owing to their unique physicochemical properties, carbon nanotubes (CNTs) are today an object of extensive fundamental and applied research [1–4]. Specifically, CNTs are very promising for emitters in field-emission electron devices [5–7], since they offer both a high electrical conductivity and a high aspect ratio (up to 1000). Accordingly, the electric field strength at the tip of a nanotube may by several hundreds of times exceed the volume-averaged field strength produced by an external source. Because of these features, the emission current may be extremely high at a comparatively low applied voltage.

Usually, the *I–V* characteristic of an individual nanotube is described well by the Fowler–Nordheim classical expression [8], which implies electron tunneling through a barrier formed by the crystal structure of the conductor and an external electric field. However, when considering the performance of CNT-based cathodes, one should take into account that the total emission current produced by a large CNT array is not equal to the arithmetic sum of the currents from individual emitters. First, a statistical parameter spread from tube to tube and the nonlinear emission current versus applied voltage dependence result in an essentially non-Fowler–Nordheim run of the *I–V* characteristic at low field strengths [9, 10].

Another reason for the difference between the *I–V* characteristics of a single nanotube and a large array is associated with the screening effect [11]. It shows up in the dependence of the field amplification factor of an emitter in the array on the emitter spacing. An increase in the field strength near the CNT tip can be characterized by the field amplification factor

$$\beta = \frac{E_{\max}}{E_{\mathrm{av}}}, \tag{1}$$

where $E_{\max}$ is the maximum field strength near the tip and E_{av} is the average field strength, which is determined as the applied voltage divided by the electrode spacing (it is assumed that electrode spacing H far exceeds nanotube height h).

Screening is absent for a single nanotube or for sparse nanotubes in the array. In this case, the field amplification factor depends on the nanotube geometry and reaches a maximal value. As the spacing shrinks, the field amplification factor decreases, approaching unity when the nanotubes merge to form a continuous flat surface (zero spacing).

Because of the screening effect, the optimal performance of CNT-based cathodes is essentially a tradeoff between a high field amplification factor, which grows with emitter spacing, and a high density of the emission current, which requires that the array be close-packed. This situation brings forth the problem of evaluating the screening effect under the operating conditions of a field array cathode and determining the optimal average emitter spacing that provides a maximal density of the emission current. The aim of the present study is to numerically tackle this problem. With this in mind, we developed a numerical approach to calculating the field strength distribution near the tip of a nanotube in arrays with different tube spacings. A method developed for solving the Laplace equation in the case of intricate-shape emitters allows for estimating the effect of screening on the emissivity and an optimal distance between nanotubes.

DESCRIPTION OF THE METHOD

The effect of screening on the emission properties of CNT emitters was mathematically simulated with the SAFANT (Simulation of Amplification Factor of NanoTubes) program, which makes it possible to calculate the spatial distribution of the field strength and potential near the top of a nanotube in a CNT array placed on a flat substrate. From the results of simulation, one can determine field amplification factor β at various average distances between the emitters.

The program solves the Laplace equation

$$\Delta\varphi = 0 \qquad (2)$$

(Δ is the Laplacian) for electric potential $\varphi(x, y, z)$ near a nanotube. The first boundary conditions for Eq. (2) correspond to the case when conducting nanotubes are placed on an infinite conducting grounded substrate normally to its surface, so that their potential is zero. The other boundary condition is potential U_0 on the counterelectrode, which also represents an infinite plate H distant from the substrate. It is assumed that electrode gap H far exceeds characteristic nanotube height h, so that the nanotubes do not affect the spatial distribution of the potential in a major part of the electrode gap. A nanotube has the form of an extended cylinder bounded by a flat top. The program allows for variation of the nanotube geometry in order to take into consideration a natural spread of the emission properties. In addition, provision is made in the model suggested to vary the arrangement of the nanotubes on the substrate so as to include the effect of mutual screening on the emission properties of neighboring CNTs. Such an approach yields a reasonable description both of a single CNT and of a large CNT array and still requires moderate computational resources. Because of the complex geometry of the problem, Laplace equation (2) is impossible to solve analytically; therefore, numerical computing techniques are needed [9].

Equation (2) with the boundary conditions specified above was solved numerically by iterations. The criterion for iteration termination has the form

$$\max(|\varphi^{(n+1)} - \varphi^{(n)}|) \le \varepsilon, \qquad (3)$$

where $\varphi^{(n)}$ and $\varphi^{(n+1)}$ are the field potentials at nth and $(n + 1)$th iterations, respectively, and ε is a desired solution (calculation) accuracy.

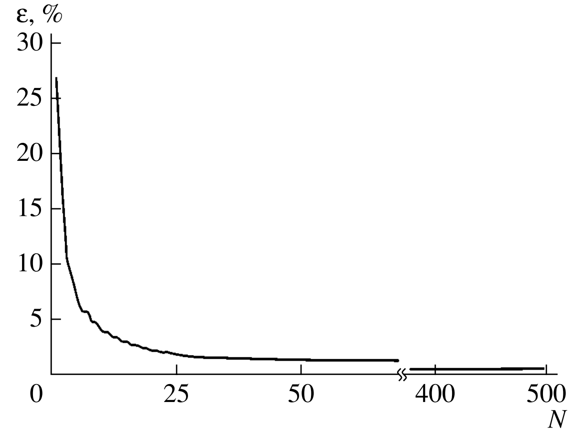


Fig. 1. Maximal error ε in calculating the field potential by the Laplace equation vs. the number of iterations N.

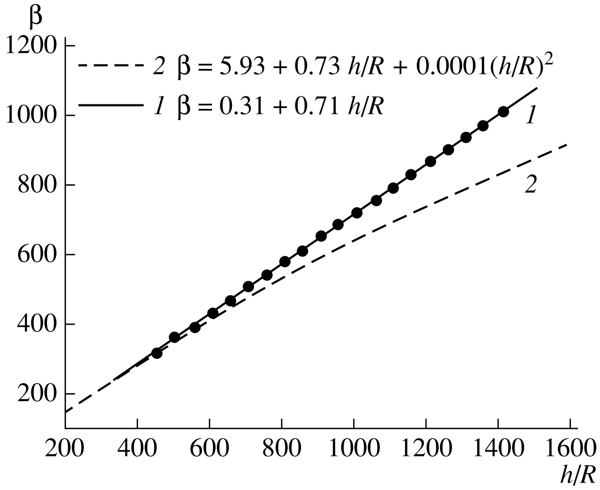


Fig. 2. Field amplification factor near the tip of a nanotube vs. the aspect ratio. Filled circles, computation with the SAFANT program; the continuous line, interpolation. The dashed line is taken from [12].

The choice of ε to a great extent depends on the number of iterations executed by the program. Figure 1, where calculation error ε is plotted versus the number of iterations, shows that the accuracy rises with the number of iterations, as was expected. An acceptable accuracy of 1–2% is attained after approximately 50 iterations, and just this number was used in the calculations. Thus, using the SAFANT program, we first find the electric field distribution near a CNT within an array. The next step is to determine the spatial distribution of the field potential as a function of a spread in the parameters of individual CNTs and of the nanotube spacing. With these distributions known, we can proceed to calculation of the emissivity of the cathode array and find an optimal (in terms of the maximal emission current density) nanotube spacing with regard to mutual screening.

RESULTS AND DISCUSSION

In order to demonstrate the potential of the program, we first performed a test computation of field amplification factor β near the tip of a single nanotube as a function of its aspect ratio $\alpha = h/R$ (R is the nanotube radius). In Fig. 2, the result obtained is compared with

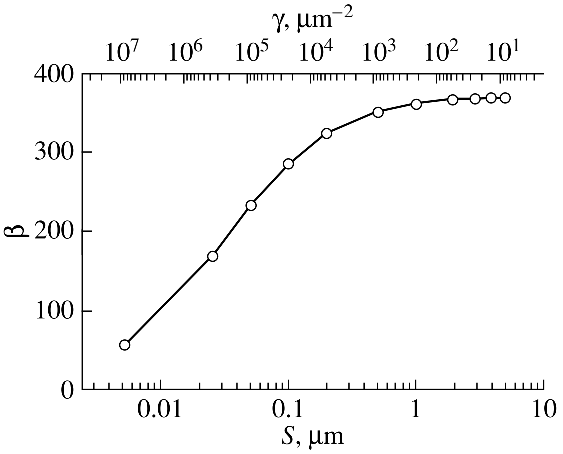


Fig. 3. Field amplification factor near the tip of a nanotube in the array vs. average nanotube spacing S and nanotube surface density γ.

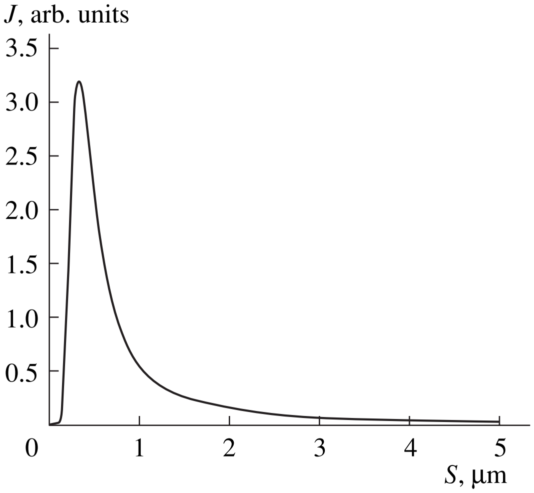


Fig. 4. Field emission current density J in the CNT cathode array vs. average spacing S, as follows from Fowler–Nordheim relationship (4) in view of screening.

the same dependence found in [12], where a nanotube was modeled by stacked identical conducting spheres. At high aspect ratios, the curves are seen to diverge, which can be explained by different approaches to modeling the nanotube structure.

The effect of mutual screening between neighboring nanotubes on the emission properties of a CNT cathode array was computed for a square array of 225 regularly arranged nanotubes, each having the form of a flat-top cylinder with diameter $d = 1$ nm and height $h = 1$ μm. The nanotubes occupy the vertices of elementary squares whose size S was varied. Figure 3 shows the computed dependence of the field amplification factor on both the emitter spacing and emitter density in the array. As the spacing diminishes, factor β decreases, tending to a minimal value, which equals unity in the limiting case of zero spacing (i.e., when the tops of the CNTs produce a continuous surface). At wide spacings ($S > h/2$), the amplification factor approaches a value typical of a single nanotube.

Emission current i from a single nanotube is described by the well-known Fowler–Nordheim law [7, 8]

$$i = C_1 E^2 \exp\left(-\frac{C_2}{E}\right), \tag{4}$$

where $E = \beta E_{\mathrm{av}}$ is the field strength near the tip and parameters C_1 and C_2 are defined by the geometry and emissivity of the emitter.

The amplification factor versus average spacing dependence found above for an array of identical nanotubes suggests that the dependence of emission current density $j \sim i/S^2$ on the average spacing is nonmonotonic, peaking at $S \sim h/2$. This dependence calculated is shown in Fig. 4. The emission current density is maximal at $S \approx 0.3h$. This value is approximately one order

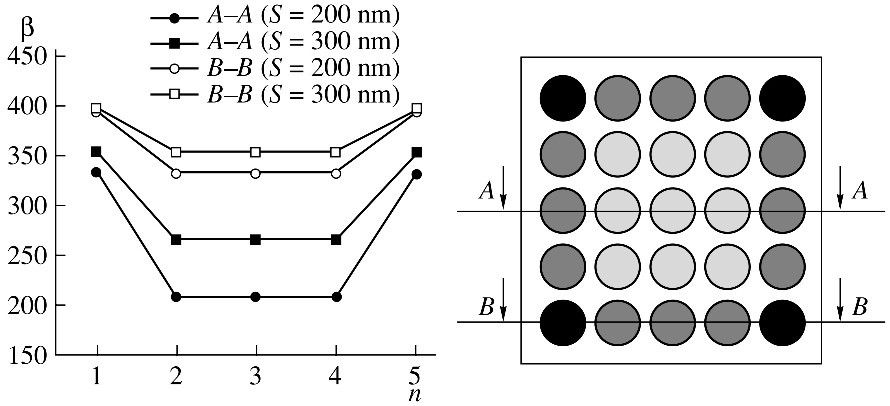


Fig. 5. Field amplification factor for CNTs entering into a 5 × 5 array vs. their position and average spacing S. n is the number of a nanotube in the row.

of magnitude lower than that obtained in [11] possibly because of different approaches to modeling the CNT structure. The peak in the current density versus the average spacing curve indicates that the emission properties of a CNT-based field cathode should be optimized by appropriately choosing the distance between the emitters.

It should be noted that the calculations were performed for a limited number of nanotubes in an array because of the need to cut the machine time. In the calculations, we considered screening for inner nanotubes alone, thereby disregarding edge effects arising at the boundaries of the array. At the boundary, screening is obviously weaker, so that outer nanotubes are bound to have a higher field amplification factor than the inner ones. To estimate edge effects, we calculated the amplification factor as a function of the position of a nanotube in an array. The results presented in Fig. 5 count in favor of this supposition: the amplification factor near the nanotubes placed at the boundaries is higher. However, the edge effects are seen to be essential for only one (extreme) row of elements and can therefore be neglected in arrays of $n \gg 1$ elements. Certainly, the edge effects become still less significant as the nanotube spacing grows, which also follows from the curves in Fig. 5.

CONCLUSIONS

Our results indicate that screening appreciably lowers the field amplification factor near carbon nanotubes constituting a field-emission cathode array if the spacing between the nanotubes in the array and their height are comparable to each other. Accordingly, the emission current density reaches a maximum at a certain spacing, which should be taken into account in cathode performance optimization.

ACKNOWLEDGMENTS

This work was supported by the International Science and Technology Center (project no. 2484), CRDF (the program “Plasma” of the Research and Education Center), and the “State Program in Support” of Leading Scientific Schools of the Russian Federation (Biberman school).

REFERENCES

1. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes* (Academic, San Diego, 1996).
2. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (World Sci., London, 1998).
3. T. W. Ebbesen, *Carbon Nanotubes: Preparation and Properties* (CRC, Boca Raton, 1997).
4. A. V. Eletskii, Usp. Fiz. Nauk **167**, 945 (1997) [Phys. Usp. **40**, 899 (1997)].
5. Yu. V. Gulyaev *et al.*, in *Proceedings of the 7th International Vacuum Microelectronics Conference, Grenoble, 1994*, p. 322; J. Vac. Sci. Technol. B **13**, 234 (1995); Yu. V. Gulyaev *et al.*, Mikroelektronika **26** (2), 84 (1997); L. A. Chernozatonskii, *et al.*, in *Proceedings of the 8th International Vacuum Microelectronics Conference, Portland, 1995*, p. 363; Chem. Phys. Lett. **233**, 63 (1995).
6. W. A. De Heer, A. Chatelain, and D. Ugarte, Science **270**, 1179 (1995).
7. A. V. Eletskii, Usp. Fiz. Nauk **172**, 401 (2002) [Phys. Usp. **45**, 369 (2002)].
8. R. Gomer, *Field Emission and Field Ionization*, 2nd ed. (AIP, New York, 1993).
9. G. S. Bocharov, A. V. Eletskii, and A. V. Korshakov, Rev. Adv. Mater. Sci. **5**, 371 (2003).
10. A. N. Obraztsov *et al.*, Diamond Relat. Mater. **12**, 446 (2003).
11. L. Nilsson *et al.*, Appl. Phys. Lett. **76**, 2071 (2000).
12. G. C. Kokkorakis, A. Modinos, and J. P. Xanthakis, J. Appl. Phys. **91**, 4580 (2002).

Translated by A. Sidorova

Technical Physics, Vol. 50, No. 7, 2005, pp. 948–950. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 131–133.
Original Russian Text Copyright © 2005 by Aksenov.

SHORT COMMUNICATIONS

Fractal Properties of Acoustic Location Responses

I. B. Aksenov
Tupolev State Technical University (KAI), ul. K. Marksa 18, Kazan, 420111 Tatarstan, Russia
e-mail: aksenov@cs5.kstu-kai.ru

Received August 4, 2003; in final form, July 19, 2004

Abstract—The fractal dimension of reflected ultrasonic signals in steel slabs with wittingly introduced defects is analyzed. The dependence of the fractal dimension on the defect size is demonstrated. The resolving power of a method used to estimate the fractal dimension of acoustic location responses and its applicability in practical flaw detection are considered. © *2005 Pleiades Publishing, Inc.*

Ultrasonic flaw detection is a mature method of nondestructive materials testing. It is stated in [1] that ultrasonic inspection has advanced to the point where new approaches are nothing else than a mere refinement of the existing ones. Categorization of irregular surfaces according to fractal dimension [2] suggests that the fractality of the reflecting boundaries of surface irregularities and defects are embodied in location signals, which can be used to advantage in practical flaw detection. The investigation of acoustic responses from the rough surfaces of corroding cylindrical tubes [3] indicates that the fractal dimension of acoustic responses varies with the roughness. For smooth surfaces, the lower limit of the Hausdorff dimension for an acoustic response is estimated as $D_2 \approx 2$, and when the peak-to-peak roughness varies from 0.2 to 6 mm, the estimate of D_2 ranges from 1.85 to 0.6. Significantly, the thickness of corrosion areas in samples with $D_2 \approx 0.6$ substantially exceeded the tube wall thickness.

Here, we report the results of fractal analysis of ultrasonic location responses from standard SO-2 samples with and without artificial defects (through holes). The samples were tested by means of ultrasonic flaw detection with sensors of two types.

In the general case, a defect has a reflecting surface in the form of a superposition of relatively large smooth asperities (from 5 to 10 wavelengths in size) and slightly rough fine irregularities (with a size much smaller than the wavelength). A location response contains a probe signal, signals reflected from defects and the walls of the object, Gaussian noise, etc. The array of location responses was used to estimate the lower boundary of the Hausdorff–Bezikovich dimension.

Figure 1 estimates the fractal dimension for the first series of location responses from the defect-free sample and from those with artificial defects of 1.5 and 5 mm in diameter. The propagation velocity of ultrasonic waves in the sample was 3200 m/s. The lower limit of the fractal dimension calculated by the correlation integral method [4] was estimated to be $D_2 = 1.37$ for the defect-free sample, $D_2 = 0.44$ for the sample with the defect of diameter 1.5 mm, and $D_2 = 0.3$ for the sample with the defect of diameter 5 mm.

To determine the applicability limits of the approach proposed more accurately and estimate its resolving power, we made additional measurements on the same samples using a Moldaviya sensor. The size of defects was varied from 0.5 to 5 mm. The propagation velocity of acoustic waves was 6000 m/s.

Figure 2 compares the fractal dimensions of location responses from defects of sizes varying from 0 to 5 mm in 1-mm step. The Hausdorff dimension is seen to decrease with increasing defect size in accordance with the general tendency. Table 1 gives the estimates of the Hausdorff dimension.

The emergence of additional reflecting boundaries in the cavity due to the presence of the defect changes the number of possible resonances in it, as observed in the location response under wide-band excitation of the cavity. The power spectra of the responses from the

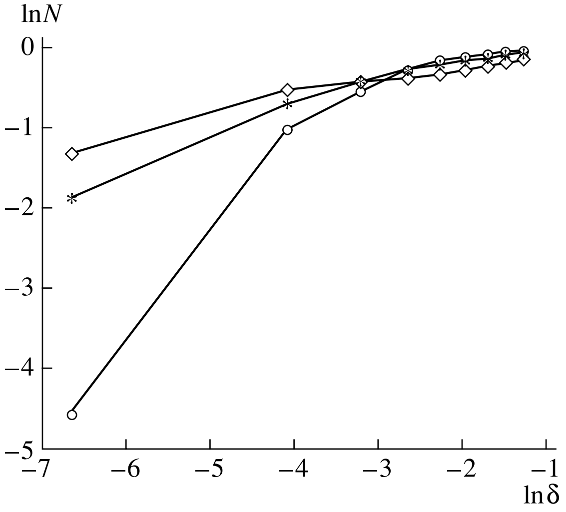


Fig. 1. Estimation of the fractal dimension for the samples from the first series of acoustic measurements. –○– $D_2d = 0$, –∗– $D_2d = 1.5$, –◇– $D_2d = 5.0$.

1063-7842/05/5007-0948$26.00 © 2005 Pleiades Publishing, Inc.

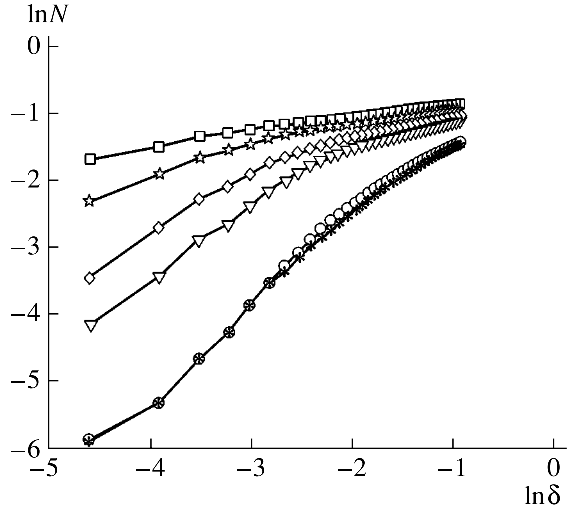


Fig. 2. Estimation of the fractal dimension for the samples from the second series of acoustic measurements. –○– $D_2d = 0$, –∗– $D_2d = 1$, –▽– $D_2d = 2$, –◇– $D_2d = 3$, –☆– $D_2d = 4$, –□– $D_2d = 5$.

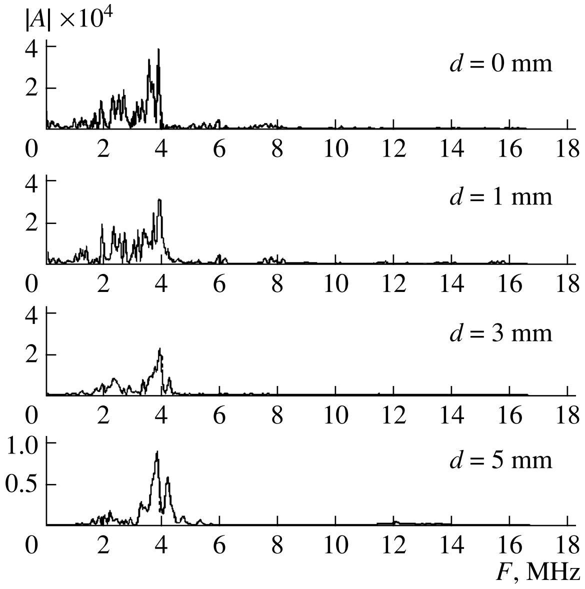


Fig. 3. Power spectrum of location responses from the samples with defects of sizes from 0 to 5 mm.

imperfect samples contain extra spectral components, while not allowing direct estimation of the defect size (Fig. 3).

The application of fractal dimension D_2 (which is also referred to as the correlation dimension) in practice is appropriate because of the power-type dependence of the power spectrum on the fractal dimension, $S(f) \approx k/f^{\beta}$ [2, 4]. Exponent β in the law of variation of the power spectrum and the surface irregularity are in one-to-one correspondence, in contrast to the Fourier estimate of the power spectrum, which gives a general variation of the location response's spectral composition.

The total number of excited modes in homogeneous cavities with smooth boundaries is proportional to the ratio of the characteristic size of the cavity (volume or surface area) to the cube and square, respectively, of the wavelength corresponding to the maximal frequency excited in the cavity [5]. In the presence of fractal boundaries, corrections used to estimate the number of modes excited contain frequency-dependent factors in lower powers. An increase in the number of resonances in the presence of fractal boundaries can be estimated using the Barry formula

$$|\Delta N(f)| = \left(L\frac{f}{c}\right)^{D}, \tag{1}$$

where ΔN is the increase in the number of resonances, L is the characteristic length, f is the frequency, c is the velocity of sound, and D is the fractal dimension.

Since the ratio f/c is the reciprocal of the wavelength corresponding to the maximal frequency excited in the cavity, the resolving power, in essence, depends on the ratio between the characteristic size of the defect and the wavelength of the cavity mode. This ratio can be written in the form

$$D = \frac{\ln(\Delta N(f))}{\ln(L/\lambda)}, \tag{2}$$

where L is the characteristic size (length) of the defect and λ is the wavelength corresponding to the maximal frequency excited in the cavity.

This relationship closely resembles the definition of the Hausdorff–Bezikovich dimension, which is formally written as the limit of the ratio between the logarithm of number N of hyperspheres covering the curve and the logarithm of the reciprocal of diameter δ of the hypersphere,

$$D = \lim_{\delta \to 0}\frac{\ln(N(\delta))}{\ln(1/\delta)}. \tag{3}$$

Following such a formalism, one can physically treat the fractal dimension for acoustic location responses as the ratio between the logarithm of the number of oscillation modes excited in the cavity and

Table 1

Defect size	D_2
0	1.47
1	1.36
2	0.84
3	0.52
4	0.32
5	0.19

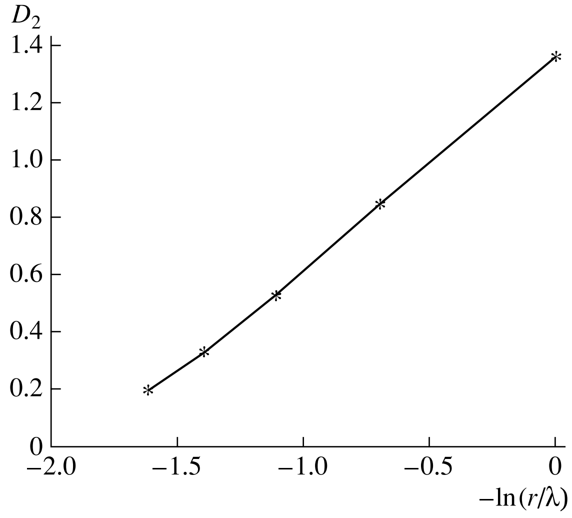


Fig. 4. Fractal dimension vs. the reciprocal logarithm of the defect relative size.

the logarithm of the reciprocal wavelength corresponding to the maximal frequency,

$$D_a \approx \frac{\ln(\Delta N(1/\lambda))}{\ln(1/\lambda)}. \tag{4}$$

In the second series of measurements, the minimal wavelength (corresponding to a maximal probe pulse repetition frequency of ≈3 MHz)) was 0.5 mm and so was the increase in the perimeter of the defect. The lower boundary (D_2) of the fractal dimension was estimated using the correlation integral. Table 2 summarizes the estimates of the Hausdorff dimension upon subsequently changing the inclination of the sensor axis by 5°.

Table 2

Defect size	D_2
0.5	1.45–1.5
1.0	1.34–1.47
1.5	0.96–1.2
2.0	0.7–1.13
2.5	0.6–0.96
3.0	0.5–0.54
3.5	0.29–0.52
4.0	0.24–0.31
4.5	0.21–0.33
5.0	0.2–0.22

As follows from the experimental results, the actual resolving power is half as large as expected.

Figure 4 shows the dependence of the estimate of D_2 on the reciprocal logarithm of the relative size of the defect, which is nearly linear. The range of measurements that is necessary for practical flaw detection is attained by varying the probe pulse repetition frequency.

It should be noted in conclusion that the method proposed is intended for cheap (hardware-nonintensive) nondestructive testers for inspecting structural elements whose properties may vary during service (due to fatigue cracks, corrosion, etc.). In hardware, the computation of the correlation integral reduces to the computation of sums in the form [5]

$$C(r) = \lim_{N\to\infty}\frac{1}{N^2}\sum_{ij} H(r - |x_i - x_j|),$$

where H is the Heaviside function, which equals unity when a pair of counts of the response that map the phase space falls into a cell of size r; otherwise, it equals zero.

For a realistic number of counts, N = 400–500, the hardware necessary for computing such sums includes devices based on readily available PIC controllers and custom programmable logic arrays.

REFERENCES

1. I. N. Ermolov, Defektoskopiya, No. 2 (2003).
2. J. Feder, *Fractals* (Plenum, New York, 1988; Mir, Moscow, 1991).
3. I. B. Aksenov and A. L. Tukmakov, *Electronic Instrument Building: A Collection of Articles* (Kazansk. Gos. Tekh. Univ., Kazan, 2001), No. 4 (20).
4. S. V. Bozhokin and D. A. Parshin, *Fractals and Multifractals* (NITs RKhD, Izhevsk, 2001) [in Russian].
5. M. Shreder, *Fractals, Chaos, and Power Laws: Miniatures from Infinite Paradise* (RKhD, Izhevsk, 2001) [in Russian].

Translated by N. Wadhwa

Technical Physics, Vol. 50, No. 7, 2005, pp. 951–953. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 134–136.
Original Russian Text Copyright © 2005 by Protasevich.

SHORT COMMUNICATIONS

RF Discharge Glow versus Air Humidity

E. T. Protasevich

Received September 13, 2004

Abstract—Based on photographs of rf discharge glows, it is concluded that the shape of the glowing object depends on the moisture content in it. © *2005 Pleiades Publishing, Inc.*

A rebirth of interest in plasmoids arising in humid air has been observed in recent years [1, 2]. Methodologically, today's experiments are virtually replicates of those performed by Plante and Weber as early as 120 years ago [3]. For example, Weber noticed that "the ball is brilliantly colored and moves near the water surface…, its color and luminosity may vary depending on the electrode material and applied voltage." This description fully coincides with the conclusions drawn in [1, 2].

In this paper, the author calls the reader's attention to the fact that the shape of the rf discharge glow depends on the moisture content in the ambient air. The experimental scheme is detailed elsewhere [4–7]. Note only that the discharge was initiated in 35-cm-long quartz tubes 7.2 cm in diameter at pressures of 133–4000 Pa. An air + H_2O mixture was ionized in an rf field at frequencies of 36–37 MHz. The generator operated in the pulsed mode (pulse width t = 10–60 ms, pulse repetition rate F = 1–5 Hz). The rf power ranged from 5 to 60 kW.

To gain insight into the physical nature of the forming plasma object, it is essential to trace the evolution of the rf discharge shape with the moisture concentration in the air + H_2O mixture. Figure 1 shows a typical discharge glow in the humid air at pressures of 133–4000 Pa. As is seen from Fig. 1, even without preliminary drying of the air, the glow occupies the entire discharge volume. The uniformity of the electron concentration in the plasma along the discharge tube depends on its length. Using two or more pairs of external ring electrodes, one can easily generate a plasma in discharge tubes of length 1 m or larger.

However, as the moisture content grows, the region of plasma glow shrinks, occupying the area near the external ring electrodes (see Fig. 2). There is an optimum concentration of H_2O molecules that provides the most effective cooling of the plasma [4]. In this case, the glow takes the form of a disk with a clear-cut boundary between the discharge and environment. In both cases, we are dealing with a quasi-stationary discharge, when the rf field ionizes the humid air for a long time (from a fraction of a second to several minutes or even hours).

When water is introduced into the discharge rapidly, the H_2O molecules, being involved in a number of complex plasma-chemical reactions, decompose [4, 8]. As a result, the plasma is partially cooled [9, 10] and a "warmer" plasma bunch separates out of the top of the discharge and floats up under the action of the buoyancy force. To record this process, a special vertical section was provided in the discharge tube. One such

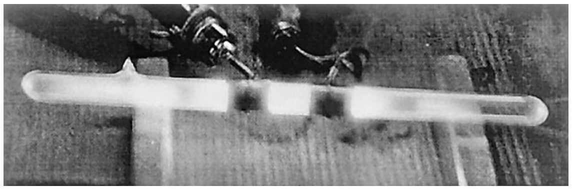


Fig. 1. RF discharge plasma glow in dry air at a pressure of about 1 kPa.

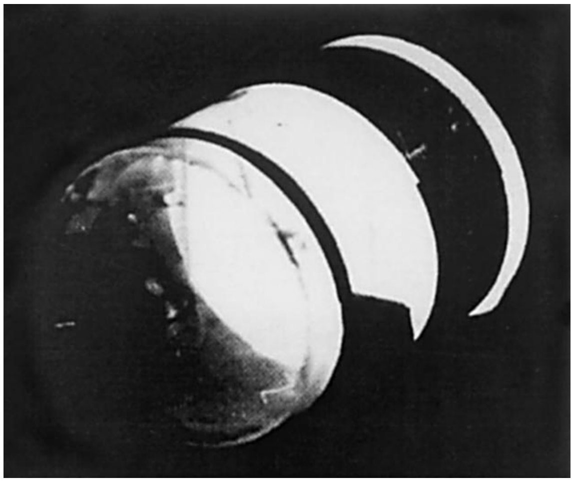


Fig. 2. Plasma disk in the rf discharge that is observed at $0.5 \times 10^{20} < N/p < 3.5 \times 10^{20}$ m^{-3} Pa^{-1}, where N is the concentration of water molecules and p is the total pressure of the unionized air + H_2O mixture. The maximal plasma cooling conditions.

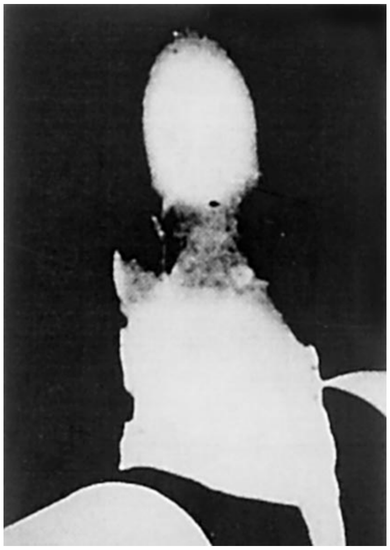


Fig. 3. Hot plasma bunch separated out from the main discharge. The moisture content in the discharge is low, $N/p < 0.1 \times 10^{20}$ m^{-3} Pa^{-1}.

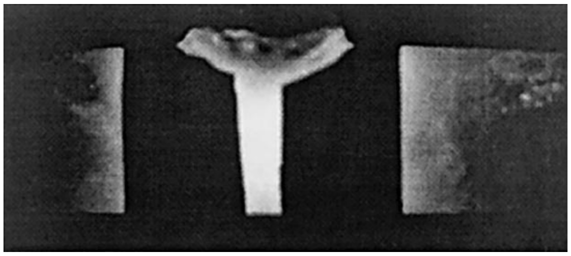


Fig. 4. RF discharge glow when $N/p > 5 \times 10^{20}$ m^{-3} Pa^{-1}. The maximum cooling of the plasma is not achieved.

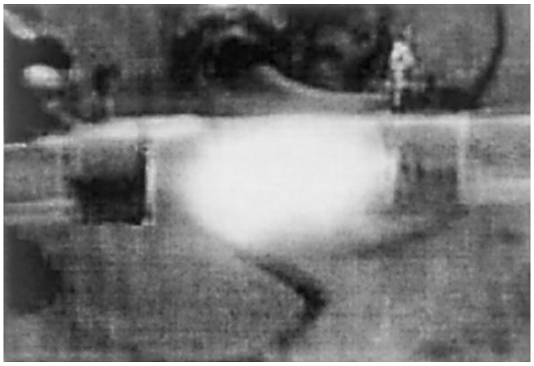


Fig. 5. Quasi-spherical bunch of the cold nonequilibrium rf discharge plasma.

bunch is depicted in Fig. 3. Its shape and size strongly depend on the concentration of water molecules in the discharge and on the form in which they enter the discharge volume. When water is introduced as an aerosol, water droplets evaporate to produce vapor. In this case, the bunch has a quasi-spherical shape similar to that shown in Fig. 3 and is easy to record with the help of filming and photography.

When the initial concentration of electrons in the plasma is $n < 10^{14}$ cm^{-3}, the discharge glow also differs from those three considered above. Because of the presence of H_2O molecules in the discharge volume, the degree of ionization of the air decreases sharply and the glow acquires the form shown in Fig. 4. Comparing all the photographs presented allows us to argue that, strictly speaking, the discharge in Fig. 4 remains a volume discharge and the average concentration of particles in it does not exceed 10^{12} cm^{-3} (as estimated using microwave interferometry in the 3-cm-wavelength range).

The fifth case seems to be the most interesting for simulating the discharge glow in humid air. It occurs when water vapor with an H_2O molecule concentration close to that meeting the plasma optimal cooling condition is introduced into the initial plasma with electron concentration $n > 10^{14}$ cm^{-3}. The associated glow is shown in Fig. 5. It is seen that, when the vapor content is close to optimal, there appears a quasi-spherical bunch of a cold nonequilibrium plasma with a clear-cut boundary between the cold plasma and environment. In Fig. 5, this bunch shows up as a quasi-spherical plasma object in the rf discharge [7]. The physical nature of such a bunch was investigated earlier in a number of works [4–10]. Unfortunately, modern equipment still gives no way of comprehensively analyzing this situation. We can state only that the density of the plasmoid is close to that of unionized air and its lifetime is 10^3–10^6 times as long as the decay time of the rf discharge plasma in the above cases (in particular, in dry air). While the decay time of the plasma in the above four cases was 10–20 μs, in the fifth case, when the quasi-spherical bunch of the cold nonequilibrium plasma formed, it was equal to approximately 1 s [7, 9].

A more detailed categorization of rf discharges at low pressures according to parameter N/p (see the caption to Fig. 2) was suggested in [10]. However, from the photos of the rf discharge presented here, it follows that the shape and properties of the glowing plasma objects strongly depend on the humidity of the atmosphere. That is why elaboration of a theory of such natural phenomena as ball lightning, bead lightning, etc., as well as efforts to reproduce them in the laboratory conditions, have failed. The key to understanding these phenomena is a better insight into the properties of rf discharges in humid air and the physicochemical processes accompanying humid air ionization.

REFERENCES

1. A. I. Egorov, S. I. Stepanov, and G. D. Shabanov, Usp. Fiz. Nauk **174**, 107 (2004) [Phys. Usp. **47**, 99 (2004)].
2. A. I. Egorov and S. I. Stepanov, Zh. Tekh. Fiz. **72** (12), 102 (2002) [Tech. Phys. **47**, 1584 (2002)].

3. J. D. Barry, *Ball Lightning and Bead Lightning: Extreme Forms of Atmospheric Electricity* (Plenum, New York, 1980; Mir, Moscow, 1983).
4. E. T. Protasevich, *Cold Non-Equilibrium Plasma: Generation, Properties, Applications* (CISP, Cambridge, 1999).
5. E. T. Protasevich, Teplofiz. Vys. Temp. **27**, 1206 (1989).
6. Yu. D. Kopytin, E. T. Protasevich, and V. I. Shishkovskii, *Effect of High-Power Laser and Microwave Radiations on the Air* (Nauka, Novosibirsk, 1992) [in Russian].
7. E. T. Protasevich, Teplofiz. Vys. Temp. **40**, 1000 (2002).
8. V. P. Grigor'ev, E. T. Protasevich, and Zh. K. Beisembaev, Sib. Fiz. Tekhn. Zh., No. 3, 57 (1992).
9. E. T. Protasevich, *New Phenomena in the Physics of Gas Discharge* (Izd. Tomsk. Politekh. Univ., Tomsk, 2002) [in Russian].
10. E. T. Protasevich and V. A. Khan, *Propagation of Electromagnetic Beams through the Atmosphere* (Izd. Tomsk. Politekh. Univ., Tomsk, 1994) [in Russian].

Translated by Yu. Vishnyakov

Technical Physics, Vol. 50, No. 7, 2005, pp. 954–956. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 137–139.
Original Russian Text Copyright © 2005 by Aĭnbinder, Vodolazov, Maksimov.

SHORT COMMUNICATIONS

Low-Field Peak Effect in Type II Superconductors

R. M. Aĭnbinder, D. Yu. Vodolazov, and I. L. Maksimov

*Lobachevsky State University,
ul. Ul'yanova 16, Nizhni Novgorod, 603950 Russia*

Received November 1, 2004

Abstract—The combined effect of a surface (edge) barrier and volume pinning on the dependence of critical current I_c on the magnetic field ($\mathbf{I} \perp \mathbf{H}_0$) in bulk type II superconductors is investigated. In low magnetic fields, there is a portion of the curve $I_c(H_0)$ where I_c grows with H_0, causing a nontrivial peak effect in this field range. Such behavior is explained by the combined effect of a surface (edge) barrier and volume pinning, the latter being rather sensitive to the transport current density distribution in a superconductor. *© 2005 Pleiades Publishing, Inc.*

The presence of a magnetic flux in the form of Abrikosov vortices in a superconductor leads to dissipative losses when the transport current passes through the superconductor. However, various mechanisms impeding the free motion of vortices are known to cause a critical current below which dissipation in the superconductor is absent. Among such mechanisms, volume pinning and the effect of a surface (edge) barrier preventing the penetration and/or escape of vortices are of greatest concern. In the majority of experiments devoted to the dependence $I_c(H_0)$, volume pinning is viewed as the only mechanism of irreversibility [1–3]. At the same time, it has been shown [4, 5] that a surface (edge) barrier has a strong effect on I_c. However, the combined effect of the most important mechanisms of irreversibility (volume pinning and surface barrier) on the dissipative properties of bulk superconductors is poorly understood. In this paper, we theoretically study the combined effect of a surface barrier and volume pinning on the field dependence of the critical current in bulk type II superconductors. A nontrivial peak effect in superconductors is predicted, which arises only in low fields as a result of competition between volume pinning and the action of a surface barrier.

Consider a plate made of a type II superconductor with bulk inhomogeneities that is placed in external magnetic field $\mathbf{H}_0 = (0, 0, H_0)$. Transport current $\mathbf{I} = (0, I, 0)$ passes through the plate. The magnetic field (and, hence, current density) distribution in bulk superconductors (with allowance for effects due to the finiteness of London penetration depth λ) is described by the London equation

$$H - \lambda^2 \frac{d^2 H}{dx^2} = n(x)\Phi_0, \qquad (1)$$

where $n(x)$ is the vortex density, $H(x)$ is the local magnetic field, and Φ_0 is the fluxon.

To take into account the edge barrier, we assume that vortices start penetrating into the superconductor when the absolute value of the current density at the edge of the plate reaches barrier-suppression (threshold) value j_s. For a real surface, j_s equals the Ginzburg–Landau depairing current density [6, 7]. Note that the current density cannot exceed the value j_s.

Let us consider a model in which depinning current density j_{dp} does not depend on H_0. It was shown [8] that, in low magnetic fields, the transition of a superconductor to the resistive state occurs in the so-called annihilation regime: vortices and antivortices penetrate into the sample from opposite sides and are annihilated inside. In this case, the current density on both sides of the superconductor equals j_s. The transition to the resistive state takes place when the regions occupied by vortices and antivortices merge together inside the specimen [9]. Such a situation persists up to some field H_1. In fields higher than H_1, the transition to the resistive state will take place in the so-called "ballistic" regime of magnetic flux creep [8]: vortices enter into the superconductor from the left and leave it from the right, generating a voltage across the plate. In the field range $H_1 < H_0 < H_2$, a superconductor with a current lower than, or equal to, the critical value has two regions free of vortices (antivortices are absent in the specimen in this regime). When the external field equals H_2, the boundary of the vortex region reaches the right-hand edge of the superconductor; hence, at fields higher than H_2 and $I = I_c$, only one vortex-free region remains in the specimen. The final expression for the critical current is

$$I_c(H_0) = \begin{cases} \dfrac{cH_{\rm dp}w}{2\pi\lambda}\sqrt{\dfrac{H_0^2 - H_s^2 - (H_{\rm dp}w/\lambda)^2}{H_0^2 - (H_pw/\lambda)^2}}, & 0 < H_0 < H_1 \\ -\dfrac{cH_0}{2\pi} + \dfrac{c}{2\pi}\sqrt{4\left(\dfrac{H_{\rm dp}w}{\lambda}\right)^2 + H_s^2}, & H_1 < H_0 < H_2 \\ \dfrac{c}{8\pi}\dfrac{H_s^2 + 4(H_{\rm dp}w/\lambda)^2 + 4H_0H_{\rm dp}w/\lambda}{H_0 + H_{\rm dp}w/\lambda}, & H_0 > H_2, \end{cases} \qquad (2)$$

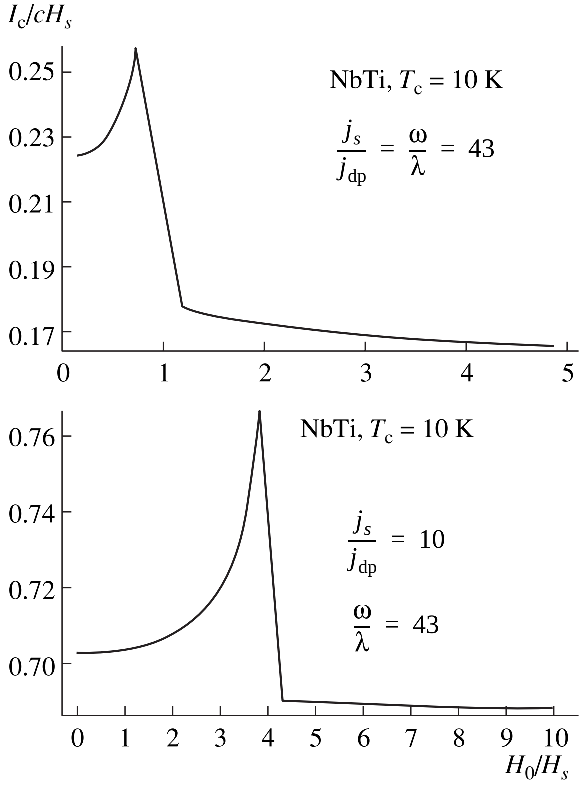
where

$$H_{\rm dp} = 4\pi\lambda j_{\rm dp}/c, \quad H_s = 4\pi\lambda j_s/c,$$

$$H_1 = -H_s/2 + H_2, \quad H_2 = \sqrt{H_s^2/4 + (H_{\rm dp}w/\lambda)^2}.$$

It follows from Eq. (2) that, in high fields ($H_0 \gg H_2$), I_c tends to a constant value equal to $cH_{\rm dp}w/(2\pi\lambda)$. This is characteristic of the Bean model (adopted in this work), in which $dj_{\rm dp}/dH_0 = 0$. Considering the real field dependence $j_{\rm dp}(H_0)$ leads to a monotonic decrease in the critical current ($dI_c/dH_0 < 0$) in high fields [10]. The dependences $I_c(H_0)$ for a NbTi low-temperature superconductor ($T_c = 10$ K) are shown in the figure. The calculations were carried out in the frame of the Bean model for different ratios $j_s/j_{\rm dp}$ and samples $w = 43\lambda$ wide. It is seen that, qualitatively, the curves behave in a similar fashion: the current first exhibits a low-field plateau; then peaks; and finally decreases slowly ($\delta I_c(H_0) \equiv I_c - I_{c\infty} \propto H_0^{-1}$), tending to a constant value ($I_c \approx I_{c\infty}$) in high fields. Note that the run of power-law tail $\delta I_c(H_0)$ at $H_0 \gg H_2$ reflects a specific character of the surface barrier [11]. As $j_{\rm dp}$ increases in the low-field limit, the relative change in I_c remains approximately constant (about 15%), while in high fields ($H_0 \gg H_2$), I_c approaches the asymptote ($I_c = cH_{\rm dp}w/2\pi\lambda$) faster. If depinning current density $j_{\rm dp}$ is much higher than barrier-suppressing current j_s, the effect of the surface barrier is negligible. For $j_s \geq j_{\rm dp}$ (volume pinning is ineffective), a surface (edge) barrier should be taken into account to adequately interpret measurements of the critical current.

The increase in I_c with increasing external magnetic field H_0 in the low-field range is a result of the combined effect of the surface barrier and volume pinning (in this situation, the transition to the resistive state takes place in the annihilation regime, when vortices and antivortices coexist; in other regimes, antivortices are absent in the superconductor). It should be noted that neither of the two mechanisms of irreversibility can singly cause an effective rise in I_c ($dI_c/dH_0 > 0$) in the low-field range ($H_0 < H_1$). In our opinion, the increase in I_c with H_0 can be explained by competition between external field H_0 and the self-magnetic field of the current in the central part of the specimen. The maximum value of derivative dI_c/dH_0 near the peak is about $0.035c$, and the maximal possible increase in the critical current (near H_1) is about 15% of its initial value.



Dependence of the critical current on the magnetic field.

At first glance, the effect being discussed could be a direct consequence of the corkscrew rule when the self-magnetic field of the transport current and the external field superpose. Numerical calculations of the critical current versus field dependence showed that the position of a maximum in this dependence differs noticeably (two to three times) from the self-field. This means that the simplified description based on the concept that the self-field of the current is the field of a maximum in the curve $I_c(H_0)$ fails. A more comprehensive theory considering the critical state of a superconductor in greater detail could allow researchers to make quantitatively correct estimates of magnetic fields H_1 and H_2.

It would be useful to compare our findings with the available experimental data for the critical current in real specimens, for example, in NbTi strips. The parameters of a surface barrier are known to be dependent largely on the sample's surface condition (for example, on the surface roughness), which is hardly controllable under real conditions. Therefore, comparison of the model parameters with the actual parameters of a superconductor will not be a trivial task.

REFERENCES

1. A. Weyers, H. Kliem, J. Lutzner, and G. Arit, J. Appl. Phys. **71**, 5089 (1992).
2. K. H. Muller, D. N. Matthews, and R. Driver, Physica C **191**, 309 (1992).
3. F. Lefloch, C. Hoffman, and O. Demolliens, Physica C **319**, 258 (1999).
4. M. Yu. Kupriyanov and K. K. Likharev, Fiz. Tverd. Tela (Leningrad) **16**, 2829 (1974) [Sov. Phys. Solid State **16**, 1835 (1975)].
5. M. Benkraouda and J. R. Clem, Phys. Rev. B **58**, 15103 (1998).
6. D. Yu. Vodolazov, I. L. Maksimov, and E. H. Brandt, Europhys. Lett. **48**, 313 (1999).
7. D. Yu. Vodolazov, I. L. Maksimov, and E. H. Brandt, Physica C **384**, 211 (2003).
8. A. A. Elistratov, D. Yu. Vodolazov, I. L. Maksimov, and J. R. Clem, Phys. Rev. B **66**, 220506(R) (2002).
9. I. L. Maksimov, Europhys. Lett. **32**, 753 (1995).
10. R. G. Mints and A. L. Rakhmanov, *Instability in Superconductors* (Nauka, Moscow, 1984) [in Russian].
11. G. M. Maksimova, N. V. Zhelezina, and I. L. Maksimov, Europhys. Lett. **53**, 639 (2001).

Translated by M. Astrov

Technical Physics, Vol. 50, No. 7, 2005, pp. 957–959. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 75, No. 7, 2005, pp. 140–142.
Original Russian Text Copyright © 2005 by Devyatisil'nyi.

SHORT COMMUNICATIONS

Three-Dimensional Gravimetry Based on the Inertial Navigation Method

A. S. Devyatisil'nyi
Institute of Automatics and Control, Far East Division,
Russian Academy of Sciences, Vladivostok, 690041 Russia
e-mail: devyatis@iacp.dvo.ru

Received November 12, 2004

Abstract—A technique with which the gravimetry problem is embedded in the problem of disposition of a three-component (three-dimensional) inertial navigation system on a stationary base is substantiated. Results of numerical simulation are presented. © *2005 Pleiades Publishing, Inc.*

(1) Measurement of inertial forces on a trajectory combined with fairly accurate model notions of the gravitational field where the motion takes place constitutes the essence of the method of trajectory determination, the so-called inertial navigation method (INM) [1, 2].

In the situation where the gravitational field is unknown but a trajectory is specified (determined by another method, e.g., by means of a satellite navigation system), the INM is well suited for determining the field. The problem thus posed is very complicated, since here we are dealing with gravimetry on a trajectory, i.e., on a mobile object.

In this work, the author considers a trajectory degenerated into a point, which greatly simplifies the problem and still retains the possibility of treating it in terms of the INM. Physically, this means turning to terrestrial navigation on a stationary base; technically, the disposition of a three-dimensional inertial navigation system (3D INS) when information about the gravitational field is incomplete. Such conditions radically distinguish both problems (gravimetry and disposition) from the conventional approaches [3, 4].

(2) The INM assumes simulation (in a given frame of reference) of equations of motion of a material point with which the moving object is identified. These equations are the Newton dynamic equations,

$$\begin{aligned} Dq &= p, \quad q(0) = q_0, \\ Dp &= G(q) + f, \quad p(0) = p_0, \end{aligned} \tag{1}$$

and the evolutionary equations for the frame of reference (the Euler–Poisson kinematic equations),

$$DA = 0, \quad A(0) = A_0. \tag{2}$$

In these equations, which are written in Hamiltonian transforms, q is the radius vector of the point on the trajectory; p is the momentum vector, which is identified with the absolute velocity vector of the point is our case; $D = d/dt + \hat{\omega}$ is the operator of absolute differentiation; ω is the absolute angular rate of rotation of the frame of reference (hereafter, we will mean only rectangular coordinate systems); $\hat{\omega}$ is the skew-symmetric matrix that is composed of the components of vector ω so that, for example, $\hat{\omega} q = \omega \times q$; $G(q)$ is the gravitational field vector; f is the nongravitational specific force vector; and A is the matrix of vector transformation from the inertial frame of reference ($o\xi = o\xi_1 o\xi_2 o\xi_3$) to a given rotating frame of reference ($ox = ox_1x_2x_3$) such that $x = A\xi$. The origin o is placed at the center of mass of the Earth. It is also assumed that coordinate trihedron ox (with axes parallel to the axes of the attending trihedron) is oriented geographically; that is, its axis ox_3 is aligned with vector q and the axis ox_2 lies in the plane of the geographic observer meridian and is directed northward.

Along with these assumptions, we should also take into account that the base is stationary, $\omega = (\omega_1, \omega_2, \omega_3)^T$, where $\omega_1 = 0$, $\omega_2 = u\cos\varphi$, $\omega_3 = u\sin\varphi$, φ is the geographic observer latitude (latitude of a gravimetric station), and u is the angular rate of rotation of the Earth.

When integrating Eqs. (1) and (2), we specify the initial conditions (q_0, p_0, A_0) and current values of ω and f, which are measured with gyros and newtonmeters (accelerometers)—inertial instruments giving the name of the method. Errors in data specified and measured cause integration errors δq, δp, and δA. In a linear approximation, the evolutionary equations for these errors can be written, in view of (1) and (2), as

$$\begin{aligned} D\delta q &= \delta p - \hat{\nu} q, \quad \delta q(0) = \delta q_0, \\ D\delta p &= \delta G(q) - \hat{\nu} p - \Delta f, \quad \delta p(0) = \delta p_0, \\ D\beta &= \nu, \quad \beta(0) = \beta_0, \end{aligned} \tag{3}$$

1063-7842/05/5007-0957$26.00 © 2005 Pleiades Publishing, Inc.

where ν and Δf are the vectors of instrumental errors from gyroscopic meters and newtonmeters, respectively; $\beta = (\beta_1, \beta_2, \beta_3)^T$ is the small-angle vector that characterizes the error of integration of the mathematical equations ($\delta A = \hat{\beta}$; and $\delta G(q)$ is the error of gravitational field strength simulation, which will be discussed below.

Note that inertial measurements and integration of Eqs. (1) and (2) are carried out relative to the axes of instrumental trihedron $oy = oy_1y_2y_3$, which is uniquely related to a measuring platform and is a physical model of trihedron ox; so, $\delta q = (\delta y_1, \delta y_2, \delta y_3)^T$.

If information about the position (platform) of the object is available from an SNS, comparing it with information from an INS generates a residual vector for the two solutions,

$$J = \delta q + \hat{\beta} q + \varepsilon, \qquad (4)$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)^T$ is the location error vector for the INS.

Let us now discuss error $\delta G(r)$ of gravitational field strength simulation. If external information on the object’s position like that considered above is absent, i.e., if the INS operates independently, $\delta G(q) = (\partial G(q)\partial q)\delta q + \Delta$ or, with regard to the dominating role of the central component in the field model,

$$\delta G(q) = -\frac{\mu}{r^3}\left(E - \frac{3qq^T}{r^2}\right)\delta q + \Delta. \qquad (5)$$

Here, μ is the terrestrial gravitational parameter, $r = |q|$, E is the unit matrix, and Δ is a gravimetric anomaly (a deviation of the field strength at the observation point from the value predicted by the model).

Note that, with $\delta G(q)$ represented in form (5), the dynamic equations of errors (the first two equations in (3)) are unstable. This fact not only limits the applicability of 3D INSs operating independently but also highlights the problem of adequacy of discrete and continuous models in light of using advanced digital computing facilities, which handle a very short time interval.

If external information on vector q from an SNS is available (the case at hand), it can be used to generate a model of field strength $G(r)$; then, instead of (5), we get

$$\delta G(q) = -\frac{\mu \delta q}{r^3} + \frac{3\mu q \varepsilon_3}{r^4} + \Delta. \qquad (6)$$

Owing to such a preliminary use of external information, the dynamic equations (and, hence, the 3D INS itself) become (nonasymptotically) stable, which facilitates subsequent numerical simulation and practical implementation of the method [4].

Completing the model considerations, we note that the set of differential and algebraic equations (3), (4), and (6) complemented by the equation $\dot{\Delta} = 0$ (which reflects the hypothesis that field local anomalies remain constant within the time of observation) is a formal statement of the inverse problem. The solution of this problem is aimed at finding the set of vectors $\{\delta q, \delta p, \beta, \Delta\}$. Thus, the gravimetry problem is embedded in a more general problem, disposition of a 3D INS on a stationary base.

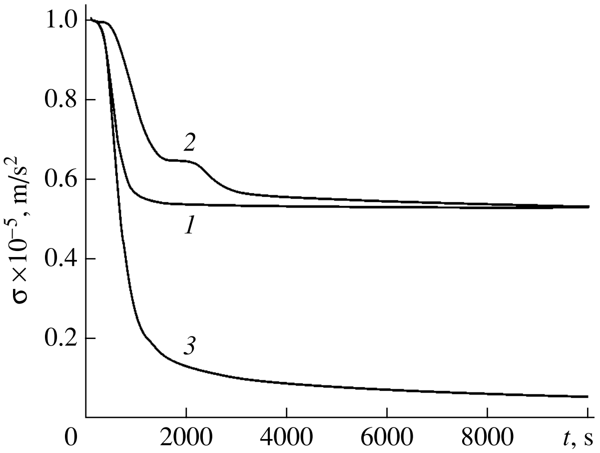
(3) To complete the formal statement of the inverse problem means to check whether it is well posed or solvable. The contemporary system theory [6] identifies the problem of solvability with the problem of observability. The essence of the latter is to establish a correspondence between the dimension of the space of images of problem’s operator and that of the declared vector of a desired solution. If a solution technique is strictly (in the Hadamard strict sense of correctness [7]) oriented to modern computing facilities and a finite accuracy of number representation in them, one should keep in mind the need for procedural support of the above correspondence (i.e., provision of the numerical stability of the solution) if it is basically established.

Analysis of the problem considered (it involves the standard procedure of constructing a basis for the space of the problem’s operator) shows that the correspondence required does exist and is violated only when vector ω of the angular rate of rotation of the Earth is aligned with the ox_3 axis, i.e., when the observation point is at one of the geographic poles of the Earth. Remarkably, the only unobservable in this case is β_3; that is, disposition of an INS cannot be accomplished in full extent (the set of vectors δq, δp, and β cannot be completely estimated) but the gravimetry problem embedded (estimation of vector Δ) can be completely solved.

(4) The consistency of the formal representations of the problem posed as applied to a specific method of its solution was verified numerically, namely, by simulating the Riccati matrix differential equation (a total of 78 equations) that describes the evolution of the solution error dispersion matrix in realizing solution procedures of the Kalman theory of optimal stochastic estimation [6].

The typical evolution of the rms estimation errors for the components of gravitational field strength vector anomaly Δ ($\sigma(\Delta_1)$, $\sigma(\Delta_2)$, and $\sigma(\Delta_3)$) is shown in the figure. The observation latitude is $\varphi = 45°$; the instrumental rms errors of inertial meters are $\sigma_f = 10^{-5}$ m/s^2 (for newtonmeters) and $\sigma_\nu = 0.0001°$/h (for gyros), respectively; and the rms location error of the SNS is $\sigma_r = 3.2$ m for each of the coordinates. It is noteworthy that the vertical component of the anomaly, Δ_3, is estimated much more accurately than the horizontal ones (Δ_1 and Δ_2). Indeed, for the steady-state values, we have $\sigma(\Delta_2)/\sigma(\Delta_1) \approx 1$ and $\sigma(\Delta_3)/\sigma(\Delta_1) \approx 0.1$.

From the steady-state rms estimation errors for the position variables ($\sigma(q_1) \approx \sigma(q_2) \approx 4$ m and $\sigma(q_3) \approx 0.4$ m), we notice that $\sigma(q_3)/\sigma(q_1) \approx 0.1$. The obvious coincidence of the error ratios for the two variables



Root-mean-square errors: (*1*) $\sigma(\Delta_1)$, (*2*) $\sigma(\Delta_2)$, and (*3*) $\sigma(\Delta_3)$.

(q and Δ) is explained by the dominating role of the central component in the gravitational field strength vector.

To complete things, the results of numerical simulation for the steady-state rms errors of estimating the kinematic (angular) variables are as follows: $\sigma(\beta_1) \approx \sigma(\beta_2) = 0.6 \times 10^{-6}$ and $\sigma(\beta_3) = 15 \times 10^{-6}$.

(5) Thus, the gravimetry method suggested in this work seems to be promising and practically feasible, especially in light of further advances in measuring and computing technologies.

REFERENCES

1. A. Yu. Ishlinskii, *Classical Mechanics and Inertial Forces* (Nauka, Moscow, 1987) [in Russian].
2. V. D. Andreev, *The Theory of Inertial Navigation: Correctable Systems* (Nauka, Moscow, 1967; Israel Program for Scientific Translations, Jerusalem, 1969).
3. *Inertial Navigation: Analysis and Design*, Ed. by C. F. O'Donnell (McGraw-Hill, New York, 1964; Nauka, Moscow, 1969).
4. *Gravitational Prospecting: A Handbook*, Ed. by E. A. Mudretsova (Nedra, Moscow, 1981) [in Russian].
5. A. S. Devyatisil'nyi and K. A. Chislov, Izv. Ross. Akad. Nauk, Ser. Teoriya i Sistemy Upravleniya, No. 5, 112 (2004).
6. R. E. Kalman, P. L. Falb, and A. A. Michael, *Topics in Mathematical System Theory* (McGraw-Hill, New York, 1969; Mir, Moscow, 1971).
7. A. N. Tikhonov and V. Ya. Arsenin, *Solutions of Ill-Posed Problems* (Nauka, Moscow, 1972; Wiley, New York, 1977).

Translated by V. Isaakyan