Technical Physics, Vol. 48, No. 2, 2003, pp. 133–137. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 1–5.

THEORETICAL AND MATHEMATICAL PHYSICS

Random Walk of a Particle in a Semiregularly Moving Plastic Medium

V. A. Antonov and A. S. Baranov

Pulkovo Astronomical Observatory, Pulkovskoe shosse 65/1, St. Petersburg, 196140 Russia

Received March 18, 2002

Abstract—A random walk of a particle is considered in a medium with randomly timed changes in the direction of its motion. An asymptotics and, in several cases, distribution of the total displacement (for finite process duration) are found. The results may be used in technological processes, geophysics, and astrophysics. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The positions of particles in various liquid and plastic mediums change when the mediums are deformed. The problem of estimating the particle displacement, together with a medium and as a result of autonomous diffusion, arises in various technological processes [1–4], as well as in geophysics and astrophysics [5–7]. One of the simple models takes place in the case when the motions of a medium are spatially ordered so that the particle displacement over a period of time is independent of its previous position. In other words, the characteristic size of inhomogeneities in the course of the deformation of a medium is substantially larger than the expected particle displacements along the coordinates x, y, and z. We will consider a more complicated case when these quantities are of the same order of magnitude elsewhere.

The displacements of a medium are assumed to be random functions of time. We will consider a simple model of random switchings (independent of the system's history) with a frequency ν. Particles are entrained by a medium with a velocity of $\mathbf{v}(v_x, v_y, v_z)$ that varies randomly and is constant between switchings. This model is referred to as semiregular. Our aim is to find the distribution law for the total displacement $\mathbf{r}(x, y, z)$ over finite time t and its asymptotics for large t.

ONE-DIMENSIONAL MODEL

First, consider one-dimensional motions along the x axis with a velocity of the same absolute value v_0 and the direction changing to the opposite at each switching. After a lapse of time t, it is necessary to find the distribution density $p(x, t)$ of the particle position in the case when it moves in the positive direction and $p(x, t)$, in the opposite direction.

As a matter of fact, this model determines a Markovian process; therefore, one might use the ordinary Kolmogorov–Feller equations [8] with a stepwise process component. However, it is convenient to use an alternate way.

Let us introduce the probability $q(x, t)dxdt$ that, in the period of time $(t, t + dt)$, both switching from the regime of leftward to rightward sliding occurs and a particle is located inside the interval $(x, x + dx)$. The function $\bar{q}(x, t)$ for the opposite switching is determined in a similar way.

Let us find the function $q(x, t)$. The number n of the last switching is not given beforehand; it is only clear that it is an even random number. If n is fixed in the definition above, we will deal with the functions $q_n(x, t)$:

$$q(x, t) = \sum_{n = 2, 4\ldots} q_n(x, t). \tag{1}$$

The quantity $q_n(x, t)dxdt$ can be considered as the probability that a two-component quantity (x, t) will belong to a two-dimensional interval $dxdt$, where t, as a random variable, performs n independent steps $t = t_1 + t_2 + t_3 + \ldots + t_n$ distributed exponentially with a density of

$$\theta_1 = \nu e^{-\nu t}. \tag{2}$$

Hence, the sum of n such steps is distributed with the density [9]

$$\theta_n = \frac{\nu^n}{(n-1)!} t^{n-1} e^{-\nu t}. \tag{3}$$

The coordinate x, by definition, is the sum of the terms with alternate signs:

$$\begin{aligned} x &= x_1 - x_2 + x_3 - \ldots - x_n \\ &= v_0(t_1 - t_2 + t_3 - \ldots - t_n). \end{aligned} \tag{4}$$

It is convenient to introduce auxiliary quantities $T = t_1 + t_3 + \ldots + t_{n-1}$ and $T' = t_2 + t_4 + \ldots + t_n$. Then,

$$t = T + T', \quad x = v_0(T - T'). \tag{5}$$

1063-7842/03/4802-0133$24.00

As is seen from the structure of T and T', they are two independent random variables each of which is distributed according to (3), where n is substituted for $n/2$. The joint density $\theta(T, T')$ is

$$\theta_n(T, T') = \frac{\nu^n}{\left[\left(\frac{n}{2}-1\right)!\right]^2} T^{n/2-1}(T')^{n/2-1} e^{-\nu(T+T')}, \quad (T, T' > 0). \tag{6}$$

In view of the mutually independent character of transformation (5), in order to perform transformation to the variables t and x, one has to make the corresponding inverse substitution in (6) and take into account the Jacobian of transformation (5), which is equal to $2v_0$. Thus,

$$q_n(x, t) = \frac{1}{2^{n-1} v_0} \frac{\nu^n}{\left[\left(\frac{n}{2}-1\right)!\right]^2} \left(t^2 - \frac{x^2}{v_0^2}\right)^{n/2-1} e^{-\nu t} \quad (t > 0, |x| \le v_0 t). \tag{7}$$

Summation over even n results in the following expression for $q(x, t)$:

$$q(x, t) = \frac{\nu^2 e^{-\nu t}}{2 v_0} I_0\left(\nu \sqrt{t^2 - \frac{x^2}{v_0^2}}\right), \tag{8}$$

where I_0 is the modified Bessel function of the first kind.

We calculate $\bar{q}(x, t)$ in a similar way. Instead of (1), we have

$$\bar{q}(x, t) = \sum_{n = 1, 3\ldots} \bar{q}_n(x, t). \tag{9}$$

The value $\bar{q}_n(x, t)dxdt$ is interpreted as the probability that a two-dimensional quantity (x, t) will belong to a two-dimensional range $dxdt$, where t, as a random variable, is given by the sum $t = t_1 + t_2 + \ldots + t_n$ and $x = x_1 - x_2 + \ldots + x_n = v_0(t_1 - t_2 + \ldots + t_n)$. We introduce auxiliary variables $T = t_1 + t_3 + \ldots + t_n$ and $T' = t_2 + t_4 + \ldots + t_{n-1}$. Formula (5) is valid in this case, and the only difference between this and the previous case is that the distributions of T and T' result from the general formula (3) by substituting n by $(n + 1)/2$ in the first case and by $(n - 1)/2$ in this case. Hence, the joint density is

$$\theta(T, T') = \frac{\nu^n}{\left(\frac{n}{2}-3\right)!\left(\frac{n}{2}-1\right)!} T^{n/2-1}(T')^{n/2-1} e^{-\nu(T+T')} \quad (T, T' > 0,\ n \ge 3),$$

$$\theta_1(T, T') = \delta(T')\nu e^{-\nu T}.$$

Having performed the transformation to the variables t and x, we find

$$\bar{q}_n(x, t) = \frac{1}{2^{n-1} v_0} \frac{\nu^n}{\left(\frac{n}{2}-3\right)!\left(\frac{n}{2}-1\right)!} \left(t + \frac{x}{v_0}\right)^{(n-1)/2} \times \left(t - \frac{x}{v_0}\right)^{(n-3)/2} e^{-\nu t} \quad (t > 0, |x| \le v_0 t, \quad n \ge 3),$$

$$\bar{q}_1(x, t) = \delta(x - v_0 t)\nu e^{-\nu t},$$

so that the summation with respect to n yields

$$\bar{q}(x, t) = \frac{\nu^2}{2 v_0} e^{-\nu t} \sqrt{\frac{v_0 t + x}{v_0 - x}} I_1\left(\nu \sqrt{t^2 - \frac{x^2}{v_0^2}}\right) + \delta(x - v_0 t)\nu e^{-\nu t}. \tag{10}$$

Now, let us turn to the determination of the functions $p(x, t)$ and $\bar{p}(x, t)$. If a particle is inside an interval $(x, x + dx)$ and is moving rightward at time t, within the time interval $(t, t + dt)$ the direction of its motion will change to the opposite with a probability of νdt. Hence, $\bar{q}(x, t)dxdt = q(x, t)\nu dtdx$, or, in other terms,

$$p(x, t) = \frac{\bar{q}(x, t)}{\nu}, \tag{11}$$

as well as

$$\bar{p}(x, t) = \frac{q(x, t)}{\nu}. \tag{11a}$$

Thus, the distributions to be found are expressed in terms of Bessel functions. Asymptotic behaviors of p and $\bar{p}$ at large t are also of interest. By analogy with the well-known central limit theorem, one may expect that the characteristic values of x will increase as $\sqrt{t}$. Using the asymptotics of the Bessel functions at large z,

$$I_0(z) \approx \frac{e^z}{2\sqrt{2\pi z}},$$

we have

$$q(x, t) \approx \frac{\nu^2}{2 v_0} \sqrt{\frac{1}{2\pi\nu}} \left(t^2 - \frac{x^2}{v_0^2}\right)^{-1/4} e^{\nu\left(\sqrt{t^2 - \frac{x^2}{v_0^2}} - t\right)}.$$

However, we have

$$t - \sqrt{t^2 - \frac{x^2}{v_0^2}} = \frac{\frac{x^2}{v_0^2}}{t + \sqrt{t^2 - \frac{x^2}{v_0^2}}} \approx \frac{x^2}{2 v_0^2 t},$$

since x^2/v_0^2 increases as t increases. In a separated factor, the second term can be neglected as well. Therefore,

$$q(x,t) \approx \frac{\nu^{3/2}}{2v_0}\sqrt{\frac{1}{2\pi t}}e^{-\frac{\nu x^2}{2v_0^2 r}},$$
$$\bar{p}(x,t) \approx \frac{1}{2v_0}\sqrt{\frac{\nu}{2\pi t}}e^{-\frac{\nu x^2}{2v_0^2 t}}. \quad (12)$$

In a similar way,

$$I_1(z) \approx \frac{e^z}{\sqrt{2\pi z}},$$
$$\bar{q}(x,t) \approx \frac{\nu^2}{2v_0}\frac{1}{\sqrt{2\pi\nu t}}e^{\nu\left(\sqrt{t^2-\frac{x^2}{v_0^2}}-t\right)} \approx q(x,t), \quad (12a)$$
$$p(x,t) \approx \bar{p}(x,t).$$

Thus, according to (12) and (12a), the asymptotic distribution of x is Gaussian, as one would expect, since the central limit theorem is generally valid for Markovian processes [8]. The influence of the initial state diminishes, although p and $\bar{p}$ do not coincide at finite t. Formulas (8) and (10), as well as (12) and (12a), can be used to estimate the particle displacement with a moving plastic medium or interstellar plasma.

GENERAL SPATIAL CASE

Consider a general spatial case when the velocity of a medium is a three-dimensional random vector $\mathbf{v}(v_x, v_y, v_z)$ with a zero mathematical expectation; variances σ_x^2, σ_y^2, and σ_z^2 with respect to principal axes; and a distribution density of $\chi(v_x, v_y, v_z)$. Let the vector $\mathbf{v}$ be chosen randomly at $t = 0$ and then experience random switchings from one value v to another independent of history with a frequency ν. The distribution of the particle displacement over time t and its asymptotics are to be found.

We will employ a previous method. Let $q(x, y, z, t)dxdydzdt$ be the probability that a switching will occur in a time interval of $(t, t + dt)$ when a particle is located in a parallelepiped with sides dx, dy, and dz around the point x, y, and z. The distribution density $p(x, y, z, t)$ is related to $q(x, y, z, t)$ by the simple relation

$$q = \nu p. \quad (13)$$

At any given number of switchings $n = 1, 2, 3, \ldots$, we have a certain probability $q_n(x, y, z, t)dxdydzdt$ and

$$q = \sum_{n=1}^{\infty} q_n. \quad (14)$$

The function q_1 is given by a distribution density for the sum of the four-dimensional quantity (x, y, z, t) where t, as a random variable, has density (2). According to our conditions, at a given $t = s$, a direct proportionality between the vectors $\mathbf{r}(x, y, z)$ and $\mathbf{v}(x, y, z)$ with the coefficient s would take place. At $n = 1$, we would have the four-dimensional distribution

$$Q_1(x, y, z, t) = \delta(t-s)\chi\left(\frac{x}{s}, \frac{y}{s}, \frac{z}{s}\right)s^{-3}. \quad (15)$$

Then, we have to average (15) over all possible s:

$$q_1(x, y, z, t) = \nu\int_0^{\infty}\frac{e^{-\nu s}}{s^3}\delta(t-s)\chi\left(\frac{x}{s}, \frac{y}{s}, \frac{z}{s}\right)ds$$
$$= \frac{\nu}{t^3}e^{-\nu t}\chi\left(\frac{x}{t}, \frac{y}{t}, \frac{z}{t}\right). \quad (16)$$

The density q_n is represented by a convolution of n distributions (16). At large n, we can directly refer to the central limit theorem. The mathematical expectations with respect to x, y, and z corresponding to q_1 are evidently equal to zero, and the variances D_x, D_y, and D_z can be calculated by (16):

$$D_x = \iiiint x^2 q_1(x, y, z, t)dxdydzdt$$
$$= \iiiint\frac{\nu}{t^3}e^{-\nu t}\chi\left(\frac{x}{t}, \frac{y}{t}, \frac{z}{t}\right)x^2 dxdydzdt$$
$$= \nu\sigma_x^2\int_0^{\infty}t^2 e^{-\nu t}dt = \frac{2\sigma_x^2}{\nu^2},$$
$$D_y = \frac{2\sigma_y^2}{\nu^2}, \quad D_z = \frac{2\sigma_z^2}{\nu^2}.$$

Now let us calculate the mathematical expectation $\langle t\rangle$ and the variance D_t with respect to t. First,

$$\langle t\rangle = \iiiint\frac{\nu}{t^2}e^{-\nu t}\chi\left(\frac{x}{t}, \frac{y}{t}, \frac{z}{t}\right)dxdydzdt$$
$$= \int_0^{\infty}\nu t e^{-\nu t}dt = \frac{1}{\nu}.$$

The second moment is

$$\langle t^2\rangle = \iiiint\frac{\nu}{t}e^{-\nu t}\chi\left(\frac{x}{t}, \frac{y}{t}, \frac{z}{t}\right)dxdydzdt$$
$$= \int_0^{\infty}\nu t^2 e^{-\nu t}dt = \frac{2}{\nu^2}.$$

Then $D_t = \langle t^2\rangle - (\langle t\rangle)^2 = 1/\nu^2$. The product moments with respect to pairs (t, x), (t, y), and (t, z) are equal to zero.

According to the central limit theorem, at large n,

$$q_n(x, y, z, t) \approx \frac{\nu^4\sqrt{2}}{\sigma_x\sigma_y\sigma_z(4\pi n)^2} \times \exp\left\{-\frac{\nu^2x^2}{4n\sigma_x^2} - \frac{\nu^2y^2}{4n\sigma_y^2} - \frac{\nu^2z^2}{4n\sigma_z^2} - \frac{\nu^2}{2n}\left(t - \frac{n}{\nu}\right)^2\right\}. \quad (17)$$

We assume a value of t to be proportional to n. It is clear that only values of n close to νt make a substantial contribution to the sum, whereas for the other values the exponent rapidly decreases. For example, it is equal to 1/2 (the peak value is 1) at $|t - n/\nu| = \sqrt{2n\ln 2}\,/\nu$ and $|n - \nu t| = \sqrt{2n\ln 2}$.

Thus, the bandwidth of values of n making a substantial contribution to the sum is proportional to n_0, where $\sqrt{n_0}$ corresponds to the middle of the band and $\sim\sqrt{\nu t}$. The values of n appearing in the other exponents, as well as the preexponential factors, do not change significantly; therefore, n can be substituted by the mean value $n_0 = \nu t$. Then, one has to sum up only the last exponents where the difference between n and n_0 can be neglected. In view of the comparatively small step of the exponent (of the order of $((\nu t)^{1/2})$ when n changes by 1 in the range under consideration, the sum

$$\sum_n e^{-\frac{1}{2\nu t}(n-\nu t)^2}$$

can be replaced by the corresponding integral

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2\nu t}(n-\nu t)^2} dn = \sqrt{2\pi\nu t}.$$

Therefore,

$$q \approx \frac{2\nu^{5/2}\sqrt{\pi}}{\sigma_x\sigma_y\sigma_z(4\pi)^2 t^{3/2}} \exp\left\{-\frac{\nu}{4t}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2}\right)\right\}$$

and hence,

$$p(x, y, z) \approx \frac{\nu^{3/2}}{8(\pi t)^{3/2}\sigma_x\sigma_y\sigma_z} \times \exp\left\{-\frac{\nu}{4t}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2}\right)\right\}. \quad (18)$$

The integral of (18) over the entire space is equal to unity, as expected. The distribution (18) is Gaussian with variances increasing as t increases.

An exact calculation of $p(x, y, z, t)$ at a finite n is somewhat difficult. They are obviated to a certain extent by using the method of characteristics. First, let us introduce such a function for the initial density χ:

$$f(\tau_x, \tau_y, \tau_z) = \iiint \chi(v_x, v_y, v_z) e^{i(v_x\tau_x + v_y\tau_y + v_z\tau_z)} dv_x dv_y dv_z. \quad (19)$$

We also define characteristics for q_n as

$$F_n(\tau_x, \tau_y, \tau_z, \tau) = \iiiint q_n(x, y, z, t) e^{i(x\tau_x + y\tau_y + z\tau_z + t\tau)} dx dy dz dt.$$

Then,

$$F_1(\tau_x, \tau_y, \tau_z, \tau) = \iiiint \frac{\nu}{t^3} e^{-\nu t} \chi\left(\frac{x}{t}, \frac{y}{t}, \frac{z}{t}\right) e^{i(x\tau_x + y\tau_y + z\tau_z + t\tau)} dx dy dz dt \quad (20)$$

$$= \nu\int_0^{\infty} f(t\tau_x, t\tau_y, t\tau_z) e^{t(i\tau - \nu)} dt$$

and according to the general rule,

$$F_n = (F_1)^n, \quad (21)$$

so that the characteristic function F for q is given by

$$F = \sum_{n=1}^{\infty} F_n = \sum_{n=1}^{\infty} F_1^n = \frac{F_1}{1 - F_1}. \quad (22)$$

PARTICULAR EXAMPLE

Consider an example when the characteristic function can be found [then, one has to perform the Fourier transform of $p(x, y, z, t)$]:

$$q_1(v_x, v_y, v_z) = \frac{4v_0^3}{\pi^2(v_x^2 + v_y^2 + v_z^2 + v_0^2)^3}. \quad (23)$$

This is a distribution with a finite variance:

$$\iiint v_q^2 q(v_x, v_y, v_z) dv_x dv_y dv_z = v_0^2;$$

the same result takes place for the other directions. Further calculations yield

$$f = e^{v_0\rho(1 + v_0\rho)},$$

where $\rho = \sqrt{\tau_x^2 + \tau_y^2 + \tau_z^2}$.

Then,

$$F_1 = \frac{\nu(2v_0\rho + \nu - i\tau)}{(v_0\rho + \nu - i\tau)^2},$$
$$\frac{F_1}{1-F_1} = \frac{\nu(2v_0\rho + \nu - i\tau)}{v_0^2\rho^2 - 2iv_0\rho\tau - i\tau(\nu - i\tau)}. \quad (24)$$

The Fourier transform with respect to τ_x, τ_y, and τ_z has to be applied to the last formula of (24). The transform with respect to τ can be easily performed if one expands the fraction in (24) into partial fractions; then we will obtain the sum of exponents with respect to t. Canceling the factor ν, we find the characteristic function

$$\frac{\frac{\nu}{2} + v_0\rho}{\sqrt{\nu^2 + 4v_0\nu\rho}}(e^{-t\xi_2} - e^{-t\xi_1}) + \frac{1}{2}(e^{-t\xi_2} + e^{-t\xi_1}), \quad (25)$$

for $p(x, y, z, t)$ (t is a parameter), where

$$\xi_{1,2} = \frac{2v_0\rho + \nu \pm \sqrt{\nu^2 + 4v_0\nu\rho}}{2}.$$

This formula can be verified if one calculates the asymptotics of (25) with respect to t for $x, y, z \sim \sqrt{t}$ and $\rho \sim 1/\sqrt{t}$. As is easily seen, formula (25) yields the characteristic function of the Gaussian law with the corresponding variance for large t.

CONCLUSIONS

In the model of random switchings under consideration, the Gaussian distribution of the total displacement is established in the course of time. This conclusion is evidently valid in the presence of autonomous diffusion of a particle: the results will be summed up according to the rule for independent random variables and the asymptotics will be Gaussian as well. The only question is how to estimate the variance; we have derived the corresponding formula (18). In several cases, it is possible to find exact displacement distributions at any positive t. In the one-dimensional case, this takes place in the model with equal velocities of displacements in the opposite directions. In the three-dimensional case, the corresponding characteristic function can be found for a particular velocity distribution (23); the results support the general calculations. In future, the particle displacements in a turbulent medium with a finite correlation distance may be studied in a similar way. Such attempts have already been performed [10] in somewhat awkward and inconvenient form.

REFERENCES

1. G. N. Abramovich, Dokl. Akad. Nauk SSSR **190**, 1052 (1970) [Sov. Phys. Dokl. **15**, 101 (1970)].
2. G. N. Abramovich and T. A. Girshovich, Dokl. Akad. Nauk SSSR **212**, 573 (1973).
3. V. N. Vigdorovich and A. S. Zherebovich, *Methods and Devices for Melt Mixing in the Processes of Crystallization Rectification of Metals and Semiconductor Materials* (Tsvetmetinformatsiya, Moscow, 1969).
4. M. A. Bernshteĭn, *Structure of Deformed Metals* (Metallurgiya, Moscow, 1977).
5. V. N. Nikolaevskiĭ, K. S. Basniev, A. T. Gorbunov, and G. A. Zotov, *Mechanics of Saturated Porous Media* (Nedra, Moscow, 1970).
6. A. E. Shreĭdegger, *Physics of Liquid Flow through Porous Media* (Gostekhizdat, Moscow, 1960).
7. A. Nicolas and J. P. Poirier, *Crystalline Plasticity and Solid State Flow in Metamorphic Rocks* (Wiley, Bristol, 1976).
8. I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes*, 2nd ed. (Nauka, Moscow, 1977; Saunders, Philadelphia, Pa., 1969).
9. Kh. B. Kordonskiĭ, *Probability Theory Application to Engineering* (Fizmatgiz, Moscow, 1963).
10. V. G. Lamburt, D. D. Sokolov, and V. N. Tutubalin, Astron. Zh. **77**, 743 (2000).

Translated by M. Fofanov

Technical Physics, Vol. 48, No. 2, 2003, pp. 138–145. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 6–12.

THEORETICAL AND MATHEMATICAL PHYSICS

Generalization of the Hecke Theorem for the Nonlinear Boltzmann Collision Integral in the Axisymmetric Case

A. Ya. Énder* and I. A. Énder**

** Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*

*** St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 199034 Russia*

Received July 16, 2002

Abstract—The properties of the nonlinear collision integral in the Boltzmann equation are studied. Expansions in spherical Hermitean polynomials are used. It was shown [1] that the nonlinear matrix elements of the collision operator are related to each other by simple expressions, which are valid for arbitrary cross sections of particle interaction. The structure of the collision operator and the properties of the matrix elements are studied for the case when the interaction potential is spherically symmetric. In this case, the linear Boltzmann operator satisfies the Hecke theorem. The generalized Hecke theorem, from which it follows that many nonlinear matrix elements vanish, is proved with recurrence relations derived. It is shown that the generalized Hecke theorem is a consequence of the ordinary Hecke theorem. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In this work, we study the structure of the collision operator using polynomial expansions. This method converts the collision integral to an interaction matrix. It has been shown [1–3] that the nonlinear elements of this matrix obey various relations, which are valid irrespective of the particle interaction potential. The application of these relations greatly simplifies the calculation of the matrix elements (MEs) at large indices.

As basis functions in which the distribution function (DF) in the Boltzmann equation is expanded, we take spherical Hermitean polynomials, which are the product of Sonine (generalized Laguerre) polynomials by spherical harmonics. The Boltzmann equation is replaced by an equivalent set of moment equations for the expansion coefficients.

The basic challenge impeding the wide application of the nonlinear method of moments is the calculation of MEs with large indices. In [2, 3], recurrence relations for the MEs in the isotropic case were found based on the invariance of the collision integral with respect to basis functions. Also, it was shown that these relations make it possible to construct the DF up to 8–10 thermal velocities with a "nearly" analytical accuracy. Similar recurrence relations for the MEs corresponding to the DF that is axisymmetric in terms of velocities were constructed in [1]. These relations allow one to express any nonlinear ME through a set of linear isotropic and nonisotropic MEs. All the relations are independent of a particle interaction model, as was mentioned above, since they are valid even if a preferential direction exists in a system, such as particles in a very high magnetic or electric field. The particles are aligned with the field and become asymmetric.

In this work, the structure of the collision operator and the properties of the MEs for nonoriented particles are studied in the case when the interaction potential is spherically symmetric, not arbitrary. Under these conditions, the interaction cross section depends only on the velocity magnitude and angle of scattering. Such potentials are the basic subject for study in the classical theory of gases.

In Section 1, we demonstrate polynomial expansions of the DF, the general form of the moment system that is axisymmetric in terms of velocities, and the recurrence relations [1] deduced for the MEs. Also, the Hecke results for the linear Boltzmann operator [4, 5] are analyzed.

In Section 2, we give a simple proof of the generalized Hecke theorem for the nonlinear axisymmetric case using the recurrence relations for the MEs. It becomes clear that the generalized Hecke theorem is a consequence of the Hecke theorem for linear MEs.

Finally, in Section 3, we consider the Kumar representation [6] of the nonlinear collision integral, which uses the Talmi transformations. The way to prove the generalized Hecke theorem with the Kumar results is outlined.

1063-7842/03/4802-0138$24.00

1. POLYNOMIAL EXPANSION OF THE AXISYMMETRIC BOLTZMANN EQUATION AND RELATIONS BETWEEN COLLISION INTEGRAL MATRIX ELEMENTS

Polynomial expansions are widely used in the kinetic theory of gases. For the linearized Boltzmann equation, such an expansion forms the basis for the well-known Chapman–Enskog method [7]. In the nonlinear case, this method was extended by Burnett [8, 9] and Grad [10]. According to Kumar [6], the most compact expansion (suggested by Burnett) is that in Hermitean spherical polynomials

$$H_{r,l,m} = S^{(r)}_{l+1/2} c^l Y^i_{lm},$$

where $S^{(r)}_{l+1/2}$ are Sonine (Laguerre) polynomials and $Y^i_{lm}(\Theta, \varphi)$ ($i = 0, 1$) are unnormalized real spherical harmonics. They are defined as follows:

$$Y^0_{lm}(\Theta, \varphi) = P^m_l(\cos\Theta)\cos m\varphi \quad (m = 0, 1, \ldots, l),$$

$$Y^{-1}_{lm}(\Theta, \varphi) = P^m_l(\cos\Theta)\sin m\varphi \quad (m = 0, 1, \ldots, l).$$

Here, $P^m_l(x)$ are associated Legendre polynomials. In the axisymmetric case ($m = 0$), spherical harmonics pass to Legendre polynomials P_l and the expansion is accomplished in functions $H_{r,l} = S^{(r)}_{l+1/2} c^l P_l$ with the Maxwellian weight

$$f(\mathbf{v}, x, t) = M(c)\sum C_{r,l}(x, t)H_{r,l}(\mathbf{c}),$$

where

$$M(c) = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-c^2}, \quad \mathbf{c} = \sqrt{\frac{m}{2\pi kT}}(\mathbf{v} - \mathbf{u}).$$

In this case, the nonlinear Boltzmann equation is replaced by a set of equations for the expansion coefficients $C_{r,l}$:

$$DC_{r,l}/Dt = \sum K^{r,l}_{r_1,l_1,r_2,l_2} C_{r_1,l_1} C_{r_2,l_2}. \tag{1}$$

Here, D/Dt is the differential operator that represents the left-hand sides of the moment equations, which were thoroughly analyzed, for example, in [9, 11].

An axisymmetric nonlinear ME has the form

$$K^{r,l}_{r_1,l_1,r_2,l_2} = \frac{1}{g_{r,l}}\int H_{r,l}\hat{I}(MH_{r_1,l_1}, MH_{r_2,l_2})d^3v,$$

$$g_{r,l} = \frac{(2r + 2l + 1)!!}{(2r)!!2^l(2l + 1)}.$$

The relations between the MEs are based on the invariance of the collision integral with respect to basis functions. For example, in the isotropic case [2, 3], the invariance with respect to the temperature T of the Maxwellian weighting function is used, while in the axisymmetric case, the invariance with respect to both T and the mean velocity u [1] is taken into consideration. The matrix of the bilinear collision operator is transformed by the formula

$$\hat{\hat{K}}' = \hat{D}\hat{\hat{K}}(\hat{D}^{-1}, \hat{D}^{-1}). \tag{2}$$

The matrix D of basis-to-basis transition can be constructed with the α–u representation [12].

Differentiating (2) with respect to T and u yields two types of relations between the MEs:

$$\begin{aligned}&\frac{l}{2l-1}K^{r,l-1}_{r_1,l_1,r_2,l_2} - \frac{(l+1)r}{2l+3}K^{r-1,l+1}_{r_1,l_1,r_2,l_2}\\ &- \frac{l_1+1}{2l_1+1}K^{r,l}_{r_1,l_1+1,r_2,l_2} + \frac{l_1(r_1+1)}{2l_1+1}K^{r,l}_{r_1+1,l_1-1,r_2,l_2}\\ &- \frac{l_2+1}{2l_2+1}K^{r,l}_{r_1,l_1,r_2,l_2+1} + \frac{l_2(r_2+1)}{2l_2+1}K^{r,l}_{r_1,l_1,r_2+1,l_2-1} = 0,\end{aligned} \tag{3}$$

$$\begin{aligned}T\frac{d}{dT}K^{r,l}_{r_1,l_1,r_2,l_2}(T) &= (R + L/2)K^{r,l}_{r_1,l_1,r_2,l_2}(T)\\ &+ rK^{r-1,l}_{r_1,l_1,r_2,l_2}(T) - (r_1+1)K^{r,l}_{r_1+1,l_1,r_2,l_2}(T)\\ &- (r_2+1)K^{r,l}_{r_1,l_1,r_2+1,l_2}(T).\end{aligned} \tag{4}$$

Here, $R = r_1 + r_2 - r$ and $L = l_1 + l_2 - l$, which are simple algebraic expressions for the case of power potentials. These relations suffice to construct all nonlinear MEs from given linear MEs of one type, $K^{r,l}_{r_1,l_1,0,0}$ or $K^{r,l}_{0,0,r_1,l_1}$, by applying the recurrence procedure.

If a system has a preferential direction, for example, in the presence of an electric or magnetic field, the particles are aligned with the field. The kinetic equation for such asymmetric particles was considered in [13], where the H theorem was also proved for the case of a preferential direction in the space. Such problems are much more difficult to solve than those considered in the conventional kinetic theory, where a preferential direction is absent and particles are assumed to be symmetric.

In the presence of a preferential direction, the scattering cross section depends on the magnitude of the relative velocity g and the angles $(\mathbf{g}_1, \hat{\mathbf{z}})$ and $(\mathbf{g}_2, \hat{\mathbf{z}})$, where $\mathbf{z}$ is the vector along the preferential direction and $\mathbf{g}_1$ and $\mathbf{g}_2$ are the vectors of the relative velocity before and after collision. In the conventional classical theory, the scattering cross section depends only on the relative velocity magnitude and angle between the vectors $\mathbf{g}_1$ and $\mathbf{g}_2$; that is,

$$\sigma = \sigma(g, \Theta). \tag{5}$$

As the kinetic theory of gases evolved, many results for the linear and linearized collision integral were obtained. Even Hilbert [14] suggested an iteration

scheme to solve the Boltzmann equation. In his scheme, the DF was represented in the form

$$f(\mathbf{v}, \mathbf{r}, t) = M(c)(1 + \varphi(\mathbf{c}))$$

and one equation with the nonlinear collision integral was transformed into a set of linear integral equations. Using hard spheres as an example, Hilbert showed that all these equations are linear integral Fredholm equations of the second kind with the symmetric and orthogonally invariant kernel $K(\mathbf{c}, \mathbf{c}_1)$. Orthogonal invariance is invariance under rotation of coordinate axes, i.e., the dependence of the kernel only on c^2, c_1^2, and the scalar product $(\mathbf{c} \cdot \mathbf{c}_1)$.

Subsequently, Hecke [4, 5] considered a linear integral equation with an orthogonally invariant kernel and proved that this operator does not remove functions of type $\psi(c)Y_{l,m}$ from the $Y_{l,m}$-related subspace; that is,

$$\hat{L}(\psi(c)Y_{l,m}) = Y_{l,m}\hat{L}^l[\psi(c)]. \qquad (6)$$

The operator $\hat{L}^l$ is independent of the subscript m, and the function $\hat{L}^l[\psi(c)]$ is isotropic in terms of velocity.

In [5], Hecke applied his theorem to the linear Boltzmann operator for the hard sphere model. This point merits detailed consideration. Following Hilbert, Hecke represented the linear collision integral in the form

$$L(\varphi) = k(c)\varphi - \int K_+(c, c_1\cos\Theta)\varphi(\mathbf{c}_1)d\mathbf{c}_1,$$

where

$$k(c) = 4\pi^2 e^{-c^2}\left(\frac{e^{-c^2}}{2} + \left(c + \frac{1}{2c}\right)\int_0^c e^{-u^2}du\right).$$

The Fourier transforms k_l^* of the kernel $K_+^* = K_+/(k(c)k(c_1))$ were defined by Hecke as follows:

$$k_l^*(c, c_1) = \int_{-1}^{1} K_+^*(c, c_1, \cos\Theta)P_l(\cos\Theta)d\cos\Theta.$$

The analytical expressions for $k_l^*(c, c_1)$ have the form

$$k_l^*(c, c_1) = \frac{e^{-(c^2 + c_1^2)}}{\sqrt{k(c)k(c_1)}}\left[\frac{8\pi}{2l+1}\frac{M^{2l+1}}{(cc_1)^{l+1}}\right.$$

$$+ \frac{4\pi M^{2l-1}}{(2l+1)(2l-1)(cc_1)^{l-1}} + \frac{4\pi M^{2l+3}}{(2l+1)(2l+3)(cc_1)^{l+1}} \qquad (7)$$

$$\left. + \frac{8\pi}{cc_1}\int_{|u|<M} ue^{u^2}E(u)P_l\left(\frac{u}{c}\right)P_l\left(\frac{u}{c_1}\right)du\right],$$

where

$$M = \min(c, c_1);\ E(u) = \int_0^u \exp(-x^2)dx.$$

Thus, the application of the Hecke theorem to the linear Boltzmann equation leads to the following result: if the distribution function $\varphi(c)$ is expanded in Legendre polynomials with expansion coefficients $\varphi_l(c)$, the Boltzmann operator in the axisymmetric case is reduced to a number of spherically symmetric integral operators

$$\hat{L}(\varphi) = \sum_l P_l(\cos\theta)\int K_l(c, c_1)\varphi_l(c_1)dc_1, \qquad (8)$$

where the kernels $K_l(c, c_1)$ are easily found through $k_l^*(c, c_1)$ (7) and Θ is the angle between the vector $\mathbf{c}$ and the z axis. It should be stressed that it is sufficient to consider only one angle because the isotropic operators $\hat{L}^l$ in (6) do not depend on the subscript m. The generalization to the three-dimensional case is straightforward.

The linear Boltzmann operator exhibits the property of orthogonal invariance with the proviso that condition (5) is satisfied. Hence, the Hecke theorem is always valid if the particles are nonoriented. Upon expanding in spherical Hermitean polynomials, Hecke theorem (6) imposes certain restrictions on the linear MEs. In the axisymmetric case, it follows from (6) that

$$K^{r,l}_{r_1,l_1,0,0} = \Lambda^{(1)}_{r,r_1,l}\delta_{l_1,l}, \qquad (9)$$

$$K^{r,l}_{0,0,r_1,l_1} = \Lambda^{(2)}_{r,r_1,l}\delta_{l_1,l}, \qquad (10)$$

that is, the MEs vanish if $l \neq l_1$. Formulas (9) and (10) state the Hecke theorem in matrix form. Hereafter, the linear elements $K^{r,l}_{r_1,l_1,0,0}(\Lambda^{(1)}_{r,r_1,l})$ and $K^{r,l}_{0,0,r_1,l_1}(\Lambda^{(2)}_{r,r_1,l})$ will be called linear elements of the first and second type, respectively.

2. PROOF OF THE GENERALIZED HECKE THEOREM FROM THE RECURRENCE RELATIONS

Let us show that, from the Hecke theorem for the linear elements, one can easily find the conditions where the nonlinear MEs vanish using recurrence relations (3) and (4). This statement will be proved in the axisymmetric case by the example of MEs of first type (9). First, we will take advantage of expression (4), which relates matrix elements with variable indices r,

r_1, and r_2 but fixed l, l_1, and l_2. Expression (4) can be recast in the form

$$(r_2+1)K^{r,l}_{r_1,l_1,r_2+1,l_2}(\tilde{T}) = -(r_1+1)K^{r,l}_{r_1+1,l_1,r_2,l_2}(\tilde{T}) - \left(R-\tilde{T}\frac{d}{d\tilde{T}}\right)K^{r,l}_{r_1,l_1,r_2,l_2}(\tilde{T}) - rK^{r-1,l}_{r_1,l_1,r_2,l_2}(\tilde{T}). \quad (11)$$

Substituting $r_2 = l_2 = 0$ into (11), we find, according to the Hecke theorem, that

$$K^{r,l}_{r_1,l_1,1,0} = G_1(r,r_1,l)\delta_{l,l_1},$$

$$G_1(r,r_1,l) = -(r_1+1)\Lambda^1_{r,r_1+1,l} - \left(R-\tilde{T}\frac{d}{d\tilde{T}}\right)\times\Lambda^{(1)}_{r,r_1,l} - r\Lambda^{(1)}_{r-1,r_1,l}.$$

Then, substituting $r_2 = 1$ and $l_2 = 0$ into (11) yields

$$K^{r,l}_{r_1,l_1,2,0} = G_2(r,r_1,l)\delta_{l,l_1},$$

where $G_2(r, r_1, l)$ is expressed through a linear combination of the functions G_1.

Clearly, by consecutively increasing r_2 in this way, one can find by induction the relationships

$$K^{r,l}_{r_1,l_1,r_2,0} = G_{r_2}(r,r_1,l)\delta_{l,l_1}; \quad (12)$$

that is, any nonlinear ME vanishes if $l \neq l_1$ irrespective of the value of r_2. Thus, it has been demonstrated that the collision operator retains the subspace related to a certain Legendre polynomial P_l not only on the linear background but also on an arbitrary spherically symmetric background.

Now we turn to equality (3), which follows from the invariance of the collision integral under shift of the velocity u and relates MEs with various l, l_1, and l_2. Let us transfer the fifth term to the left (i.e., we determine the ME $K^{r,l}_{r_1,l_1,r_2,l_2+1} = K^{r,l}_{r_1,l_1,r_2,l_2'}$) and take indices $l_1 \geq l_2'$. Putting first $l_2' = 1$ ($l_2 = 0$), we have

$$-K^{r,l}_{r_1,l_1,r_2,1} = \beta(l-1)K^{r,l-1}_{r_1,l_1,r_2,0} + \gamma(r-1,l+1)\times K^{r-1,l+1}_{r_1,l_1,r_2,0} - \beta(l_1)K^{r,l}_{r_1,l_1+1,r_2,0} - \gamma(r_1,l_1)K^{r,l}_{r_1+1,l_1-1,r_2,0}. \quad (13)$$

With $l = l_1 - 1$ substituted into (13), the first generalization of the Hecke theorem (see (12)) yields that the first and third terms on the right vanish and only the second and fourth terms can be other than zero. With $l = l_1 + 1$, only the first and third terms on the right can be nonzero. For all other values of l, all the terms on the right are zero according to (12). Thus, the MEs $K^{r,l}_{r_1,l_1,r_2,1}$ can be other than zero only in two cases:

$$l = l_1 - 1, \quad l = l_1 + 1. \quad (14)$$

Now we put $l_2' = 2$ in (3). Then, the following MEs are present on the right: $K^{r,l-1}_{r_1,l_1,r_2,1}$, $K^{r-1,l+1}_{r_1,l_1,r_2,1}$, $K^{r,l}_{r_1,l_1+1,r_2,1}$, $K^{r,l}_{r_1+1,l_1-1,r_2,1}$, and $K^{r,l}_{r_1,l_1,r_2+1,0}$. Applying inequality (14) to the first four of them, we find that the first and third MEs are other than zero only if $l = l_1$ or $l = l_1 + 2$, while the second and fourth are nonzero only if $l = l_1 - 2$ or $l = l_1$. The fifth ME can be nonzero only if $l = l_1$ according to (12). Combining these conditions, we come to the conclusion that the ME $K^{r,l}_{r_1,l_1,r_2,2}$ is nonzero only if

$$l_1 - 2 \leq l \leq l_1 + 2, \quad (-1)^l = (-1)^{l_1+2}. \quad (15)$$

The second condition in (15) means that the parity of l coincides with that of $l_1 + 2$. At $l_2' = 3$, conditions for l are found from (14) and (15). Increasing l_2 still further, we can easily find by induction that, according to the Hecke theorem for linear elements of the first type (see (9)), the MEs $K^{r,l}_{r_1,l_1,r_2,l_2}$ for $l_1 \geq l_2$ are other than zero only if

$$l_1 - l_2 \leq l \leq l_1 + l_2, \quad (-1)^l = (-1)^{l_1+l_2}. \quad (16)$$

In the case of linear elements of second type (10), the generalization of the Hecke theorem is performed in a similar way. First, taking the fourth term in (4) as the desired one, we arrive at the relationship

$$K^{r,l}_{r_1,0,r_2,l_2} = F(r,r_1,r_2,l)\delta_{l,l_2}.$$

Then, one can apply all the considerations concerning the zeros of MEs if the third, rather than the fifth, element in (3) is determined. In this case, it is easy to check that the MEs $K^{r,l}_{r_1,l_1,r_2,l_2}$ at $l_1 \leq l_2$ can be nonzero only if

$$l_2 - l_1 \leq l \leq l_1 + l_2, \quad (-1)^l = (-1)^{l_1+l_2}. \quad (17)$$

Combining (16) and (17) yields

$$|l_2 - l_1| \leq l \leq l_1 + l_2, \quad (-1)^l = (-1)^{l_1+l_2}. \quad (18)$$

The properties of MEs obeying (18) state the generalized Hecke theorem (GHT) for nonlinear MEs.

Thus, from the Hecke theorem for linear elements, one proves the GHT for nonlinear MEs with (3) and (4). It should be noted that the vanishing of MEs that do not satisfy the GHT conditions at $l_1 > l_2$ is a consequence of the Hecke theorem for linear elements of the first type, the vanishing of the same MEs at $l_1 < l_2$ is a consequence of the Hecke theorem for linear elements of the second type, and the presence of the corresponding zeros at $l_1 = l_2$ is proved independently for linear MEs of both the first and second types.

To conclude this section, we generalize the Hecke results for the expansion of the kernel of linear Boltzmann operator (8) to the nonlinear case. This becomes possible since recurrence relations (3) and (4) allow the

calculation of nonlinear MEs $K^{r,l}_{r_1,l_1,r_2,l_2}$ with large indices. Indeed, since the set of spherical Hermitean polynomials is complete in our case, we can write the kernel of the bilinear operator in the form

$$K(\mathbf{c},\mathbf{c}_1,\mathbf{c}_2) = M(c)\sum_{l_1,l_2}\sum_{l=|l_1-l_2|}^{l_1+l_2}\sum_{r,r_1,r_2}\frac{1}{g_{r_1,l_1}g_{r_2,l_2}} \times c^l S^r_{l+1/2}(c)P_l(x)K^{r,l}_{r_1,l_1,r_2,l_2}c_1^{l_1}S^{r_1}_{l_1+1/2}(c_1) \times P_{l_1}(x_1)c_2^{l_2}S^{r_2}_{l_2+1/2}(c_2)P_{l_2}(x_2). \quad (19)$$

Here, $x = \cos\Theta$, $x_1 = \cos\Theta_1$, and $x_2 = \cos\Theta_2$. The angles Θ, Θ_1, and Θ_2 are those between the velocity vectors $\mathbf{c}$, $\mathbf{c}_1$, and $\mathbf{c}_2$, respectively, and the z axis. The upper limit of summation over l in (19) is set according to the GHT.

If the DF is expanded in Legendre polynomials with expansion coefficients $f_l(c)$, expression (8) has the generalization

$$\hat{I}(f,f) = \sum_{l_1,l_2}\sum_{l=|l_1-l_2|}^{l_1+l_2}P_l(x)\int_0^\infty\int_0^\infty K^l_{l_1,l_2} f_{l_1}(c_1)f_{l_2}(c_2)dc_1dc_2.$$

Here, $\hat{I}(f,f)$ is the nonlinear Boltzmann collision integral. The kernels $K^l_{l_1,l_2}$ in this case are given by

$$K^l_{l_1,l_2} = M(c)c^l\sum_{r,r_1,r_2}\frac{1}{\sigma_{r_1,l_1}\sigma_{r_2,l_2}}S^r_{l+1/2}(c)K^{r,l}_{r_1,l_1,r_2,l_2}c_1^{l_1} \times S^{r_1}_{l_1+1/2}(c_1)c_2^{l_2}S^{r_2}_{l_2+1/2}(c_2),$$

where σ_{r_1,l_1} and σ_{r_2,l_2} are normalizing factors for the corresponding Sonine polynomials.

3. DERIVATION OF FORMULAS FOR MATRIX ELEMENTS AND THE PROOF OF THE GENERALIZED HECKE THEOREM BASED ON KUMAR'S RESULTS

Kumar [6] considered the moment representation of the collision integral upon expanding the DF in complex spherical Hermitean polynomials:

$$\phi^{[n,l]}_m(\mathbf{c}) = \sqrt{\sigma_{n,l}}c^l S^{(n)}_{l+1/2}(c^2)\mathcal{Y}^{[l]}_m(\Theta,\varphi),$$

where

$$\mathcal{Y}^{[l]}_m(\Theta,\varphi) = i^l(-1)^{(m+|m|)}\left[\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}\right]^{1/2} \times P_l^{|m|}(\cos\Theta)e^{im\varphi},$$

$$P_l^{|m|}(\cos\Theta) = \frac{(-1)^l}{2^l l!}(\sin\Theta)^{|m|}\frac{d^{l+|m|}}{(d\cos\Theta)^{l+|m|}}(1-\cos^2\Theta)^l.$$

He noted that these polynomials are the eigenfunctions of a quantum harmonic oscillator and showed that the nonlinear MEs can be conveniently calculated with the help of transformations first introduced by Talmi [15] in the context of the shell model of the nucleus.

Following Kumar and using for the most part his notation, we write the expansion of the collision integral in the form

$$I(\mathbf{c}_1) = w(\beta_1,c_1)\sum_{n_3,l_3,m_3}I^{(n_3,l_3)}_{m_3}(\beta_1)\phi^{[n_3,l_3]}_{m_3}(\beta_1\mathbf{c}_1),$$

where

$$I^{(n_3,l_3)}_{m_3}(\beta_1) = \int I(\mathbf{c}_1)\phi^{(n_3,l_3)}_{m_3}(\beta_1\mathbf{c}_1)d\mathbf{c}_1 = \iint(f_1'f_2' - f_1f_2)\phi^{(n_3l_3)}_{m_3}(\beta_1\mathbf{c}_1)g\sigma(g,\Theta)d\hat{\mathbf{g}}'d\mathbf{c}_1d\mathbf{c}_2;$$

$$w(\beta,c) = \left(\frac{\beta^2}{2\pi}\right)^{3/2}\exp\left(-\frac{\beta^2c^2}{2}\right); \quad \int w(\beta,c)d\mathbf{c} = 1.$$

The product of two spherical Hermitean polynomials in particle velocity before (or after) collision is expressed through the sum of the products of spherical Hermitean polynomials in center-of-mass velocity by the relative velocity of colliding (bouncing apart) particles. The expansion coefficients in this case are the Talmi coefficients

$$\phi^{[n_1,l_1]}_{m_1}(\beta_1\mathbf{c}_1)\phi^{[n_2,l_2]}_{m_2}(\beta_2\mathbf{c}_2) = \sum_{\substack{N,L,M\\n,l,m}}\left(\begin{matrix}NLM\\nlm\end{matrix}\middle|\begin{matrix}(\beta_1)n_1l_1m_1\\(\beta_2)n_2l_2m_2\end{matrix}\right)\phi^{[N,L]}_M(\Gamma\mathbf{G})\phi^{[n,l]}_m(\gamma\mathbf{g}).$$

Here,

$$\beta_1^2 = \frac{m_a}{kT}, \quad \beta_2^2 = \frac{m_b}{kT},$$

$$\Gamma^2 = \frac{m_a+m_b}{kT}, \quad \gamma^2 = \frac{m_am_b}{(m_a+m_b)kT},$$

$\mathbf{G}$ is the center-of-mass velocity, $\mathbf{g}$ is the relative velocity, and m_a and m_b are the particle masses. Then, an arbitrary expansion coefficient of the collision integral $I^{(n_3,l_3)}_{m_3}$ is represented as the sum of the products of the integrals of the center-of-mass velocity and the integrals of the relative velocities, the Talmi coefficients

entering only into the former. The integrals of the relative velocities have the form

$$R^{n,l,m}_{n',l',m'} = \int w(\gamma, g)[\phi^{[n,l]}_m(\gamma\mathbf{g}') - \phi^{[n,l]}_m(\gamma\mathbf{g})] \times \phi^{(n',l')}_{m'}(\gamma\mathbf{g}) g\sigma(g,\Theta) d\hat{\mathbf{g}}' d\mathbf{g}. \quad (20)$$

Kumar also showed that if the particles are nonoriented (he analyzed only this case), three of the five integrals in (20) can be taken and

$$R^{n,l,m}_{n',l',m'} = V^l_{n,n'}\delta_{l,l'}\delta_{m,m'}, \quad (21)$$

where

$$V^l_{n,n'} = \int_0^\infty w(\gamma, g) R_{n,l}(\gamma g) R_{n',l}(\gamma g)(\sigma_l - \sigma_0) g^3 dg,$$

$$\sigma_l(g) = \frac{(2l+1)^2}{8\pi}\int \sigma(g,\Theta) P_l(\cos\Theta) d(\cos\Theta),$$

and $R_{n,l}(\gamma g)$ is a radial function from the class of normalized Laguerre (Sonine) polynomials.

The Kronecker symbols $\delta_{l,l'}\delta_{m,m'}$ in (21) appear because of the spherical symmetry of the interaction potential. As a result, the expression for $I^{(n_3,l_3)}_{m_3}(\beta_1)$ is substantially simplified because it becomes shorter by two summations.

From Kumar's expression for $I^{(n_3,l_3)}_{m_3}(\beta_1)$, it is easy to imagine the form of the MEs in the case of expansion in normalized spherical Hermitean polynomials; namely,

$$\bar{E}^{n_3,l_3,m_3}_{n_1,l_1,m_1,n_2,l_2,m_2} = \sum_{\substack{N,L,M \\ n,n',l,m}} \begin{pmatrix} NLM \\ n'lm \end{pmatrix} \begin{vmatrix} (\beta_1) n_3 l_3 m_3 \\ (\beta_2) 000 \end{vmatrix}^* \times \begin{pmatrix} NLM \\ nlm \end{pmatrix} \begin{vmatrix} (\beta_1) n_1 l_1 m_1 \\ (\beta_2) n_2 l_2 m_2 \end{vmatrix} V^l_{n,n'}. \quad (22)$$

Here, instead of K the symbol E is used to distinguish MEs expanded in complex and real spherical harmonics.

Talmi coefficients are other than zero if certain relationships between indices are valid. Specifically, the parity conservation law must be fulfilled; that is, the Talmi coefficient

$$\begin{pmatrix} NLM \\ nlm \end{pmatrix} \begin{vmatrix} (\beta_1) n_1 l_1 m_1 \\ (\beta_2) n_2 l_2 m_2 \end{vmatrix}$$

is other than zero only if the parity of $L + l$ coincides with that of $l_1 + l_2$. If so, from (22) it follows that the ME $\bar{E}^{n_3,l_3,m_3}_{n_1,l_1,m_1,n_2,l_2,m_2}$ can be other than zero only if the parity of $l_1 + l_2$ coincides with the parity of l_3:

$$(-1)^{l_1+l_2} = (-1)^{l^3}. \quad (23)$$

Further mathematics is based on the Smirnov formula [16] that expresses Talmi coefficients through well-known Clebsch–Gordan coefficients $C^{c\gamma}_{a\alpha b\beta}$:

$$\begin{pmatrix} NLM \\ nlm \end{pmatrix} \begin{vmatrix} (\beta_1) n_1 l_1 m_1 \\ (\beta_2) n_2 l_2 m_2 \end{vmatrix} = \sum_\lambda B_\lambda C^{\lambda M_0}_{l_1 m_1 l_2 m_2} C^{\lambda M_0}_{LMlm}. \quad (24)$$

It is essential here that the coefficients B_λ do not depend on the projections of the moments of m_1, m_2, M, and m. The properties of the Clebsch–Gordan coefficients can be found, for example, in [17]. One of them that corresponds to the law of conservation of angular momentum projection implies that the Talmi coefficients can be other than zero only if

$$M_0 = m_1 + m_2 = M + m.$$

To transform the first of the Talmi coefficients in (22), we will take advantage of the following properties of the Clebsch–Gordan coefficients:

$$C^{\lambda M_0}_{l_3 m_3 00} = \delta_{\lambda,l_3}\delta_{M_0,m_3}, \quad (25)$$

$$\sum_{M,n} C^{\lambda M_0}_{LMlm} C^{l_3 m_3}_{LMlm} = \delta_{\lambda,l_3}\delta_{M_0,m_3}. \quad (26)$$

From (24), we have

$$\begin{pmatrix} NLM \\ n'lm \end{pmatrix} \begin{vmatrix} (\beta_1) n_3 l_3 m_3 \\ (\beta_2) 000 \end{vmatrix}^* = \sum_\lambda B^*_\lambda C^{\lambda M_0}_{l_3 m_3 00} C^{\lambda M_0}_{LMlm} = B^*_{l_3}(N, n', n_3, 0, L, l, l_3, 0) C^{l_3 m_3}_{LMlm}.$$

Here, the arguments of the function $B^*_{l_3}$ are explicitly substituted into formula (25). From (22) and (24), we come to

$$\bar{E}^{n_3,l_3,m_3}_{n_1,l_1,m_1,n_2,l_2,m_2} = \sum V^l_{n,n'} B^*_{l_3}(N, n', n_3, 0, L, l, l_3, 0) \times \sum_\lambda B_\lambda(N, n, n_1, n_2, L, l, l_1, l_2) C^{\lambda M_0}_{l_1 m_1 l_2 m_2} \sum_{M,m} C^{\lambda M_0}_{LMlm} C^{l_3 m_3}_{LMlm}. \quad (27)$$

Applying property (26) to the last sum, one can simplify expression (27) to

$$\bar{E}^{n_3,l_3,m_3}_{n_1,l_1,m_1,n_2,l_2,m_2} = C^{l_3 m_3}_{l_1 m_1 l_2 m_2} \sum V^l_{n,n'} B^*_{l_3}(N, n', n_3, 0, L, l, l_3, 0) \times B_{l_3}(N, n, n_1, n_2, L, l, l_1, l_2). \quad (28)$$

Thus, a nonlinear matrix element is proportional to a Clebsch–Gordan coefficient with the same indices. Note that the expression under the summation sign does not depend on the indices m_1, m_2, and m_3.

Formula (28) is one more approach to directly calculating the MEs. Like all other direct formulas, it

includes many summations, is extremely tedious, and thus, inappropriate for calculating MEs with large indices.

Recall that initially the Talmi coefficients were used in quantum-mechanical calculations, specifically, in the calculation of MEs within the shell model of the nucleus. For quantum calculations with large indices, one can invoke an analogue of the recurrence procedure developed for the kinetic theory in this and earlier works of the authors. Note that the proportionality of nonlinear MEs to the Clebsch–Gordan coefficients was proved [17] based on the invariance of the bilinear Boltzmann operator under rotation group *SO* (see (3)) (such a procedure is valid only for centrosymmetric potentials).

Let us turn back to formula (28). Owing to the fact that the expression under the summation sign is independent of the indices m, m_1, and m_2, any three-dimensional ME can be expressed through an axisymmetric one:

$$\bar{E}^{n_3, l_3, m_3}_{n_1, l_1, m_1, n_2, l_2, m_2} = \frac{C^{l_3 m_3}_{l_1 m_1 l_2 m_2}}{C^{l_3 0}_{l_2 0 l_2 0}} \bar{E}^{n_3, l_3, 0}_{n_1, l_1, 0, n_2, l_2, 0}.$$

It is obvious that in the axisymmetric case all indices m vanish, complex and real harmonics coincide, and spherical harmonics become Legendre polynomials. Also, E with zero m, m_1, and m_2 coincides with $K^{n_3, l_3}_{n_1, l_1, n_2, l_2}$ up to the normalizing factor.

It is well known [18] that, according to the rule of vector summation, the Clebsch–Gordan coefficients are other than zero only if they meet the "triangle" conditions:

$$|l_1 - l_2| \le l_3 \le l_1 + l_2. \quad (29)$$

Hence, three-dimensional MEs will be nonzero only if the triangle conditions are met. In the axisymmetric case, which is of interest to us, property (29), together with parity conservation law (23), states the GHT (see (18)).

In the particular case of Maxwellian molecules, MEs were considered in [19–21]. In the first two works cited, where the Fourier transformation is applied to the Boltzmann equation, formulas for directly calculating nonlinear MEs were derived. In [21], formulas for calculating axisymmetric nonlinear MEs are derived with the α–u representation, calculations for very large indices are performed, and tables of elements calculated are given.

It is of interest that, in the case of Maxwellian molecules, not only those MEs vanish to which the GHT is applied but also those for which the conditions

$$l = 0, \quad l_1 = l_2 = 2n + 1, \quad n = 0, 1, 2, \ldots$$

are fulfilled [22].

CONCLUSION

It is important that the rotation group and the orthogonal invariance of the kernel are used only when the GHT is stated for linear MEs. When the GHT is proved for nonlinear MEs, relationships (3) and (4) are used. These relationships are universal and follow from the invariance of the collision integral in terms of basis. When the basis is changed, only the temperature and velocity (shift) of the weighting Maxwellian change.

The GHT not only places limits on the range of nonzero MEs but also allows one to establish additional relations between MEs. From the GHT and universal recurrence relations deduced in this paper, it follows that any linear axisymmetric ME $K^{r, l}_{r_1, l_1, r_2, l_2}$ can be uniquely determined from given isotropic linear MEs $K^{r, 0}_{r_1, 0, 0, 0}$ or $K^{r, 0}_{0, 0, r_2, 0}$ [22]. For these linear elements, simple analytical formulas were given in [2]. An algorithm allowing the simple calculation of any linear axisymmetric ME was developed in [22]. Such MEs are proportional to integral brackets [23], with which transport coefficients in the classical kinetic theory of gases are calculated. Thus, the GHT can ensure a great step forward in the transport theory.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (grant nos. 00-02-16882 and 01-02-17903) and the Federal Program "Integration" (grant no. AO143).

REFERENCES

1. A. Ya. Énder and I. A. Énder, Zh. Tekh. Fiz. **72** (5), 1 (2002) [Tech. Phys. **47**, 513 (2002)].
2. A. Ya. Énder and I. A. Énder, Zh. Tekh. Fiz. **69** (6), 22 (1999) [Tech. Phys. **44**, 628 (1999)].
3. A. Yu. Ender and I. A. Ender, Phys. Fluids **11**, 2720 (1999).
4. E. Hecke, Math. Ann. **78**, 398 (1917).
5. E. Hecke, Math. Z. **12**, 274 (1922).
6. K. Kumar, Ann. Phys. (N.Y.) **37**, 113 (1966).
7. S. Chapman and T. G. Cowling, *Mathematical Theory of Non-Uniform Gases*, 2nd ed. (Cambridge Univ. Press, Cambridge, 1952; Inostrannaya Literatura, Moscow, 1960).
8. D. Burnett, Proc. London Math. Soc. **39**, 385 (1935).
9. D. Burnett, Proc. London Math. Soc. **40** (1935).
10. H. Grad, Commun. Pure Appl. Math. **2**, 311 (1949).
11. L. Sirovich, Phys. Fluids **6** (1), 10 (1963).
12. I. A. Énder and A. Ya. Énder, Dokl. Akad. Nauk SSSR **193** (1), 61 (1970) [Sov. Phys. Dokl. **15**, 633 (1971)]; Phys. Fluids **6** (1), 10 (1963).
13. A. Ya. Énder, Zh. Tekh. Fiz. **62** (1), 20 (1992) [Sov. Phys. Tech. Phys. **37**, 9 (1992)].

14. D. Hilbert, Math. Ann. **72**, 562 (1912).
15. I. Talmi, Helv. Phys. Acta **25**, 185 (1952).
16. Yu. F. Smirnov, Nucl. Phys. **27**, 177 (1961).
17. V. V. Vedenyapin, Preprint No. 38, IPM AN SSSR (Inst. of Applied Mathematics, USSR Academy of Sciences, 1981).
18. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
19. E. M. Hendrics and T. M. Nieuwenhuizen, J. Stat. Phys. **29** (3), 591 (1982).
20. V. V. Vedenyapin, Dokl. Akad. Nauk SSSR **256**, 338 (1981) [Sov. Phys. Dokl. **26**, 26 (1981)].
21. A. Ya. Énder and I. A. Énder, *Aerodynamics of Rarefied Gases* (Leningrad, 1983), pp. 197–215.
22. A. Ya. Énder, I. A. Énder, and M. B. Lyutenko, Preprint No. 1748, FTI RAN (Physicotechnical Inst., Russian Academy of Sciences, St. Petersburg, 2000).
23. J. H. Ferziger and H. G. Kaper, *Mathematical Theory of Transport in Gases* (North-Holland, Amsterdam, 1972; Mir, Moscow, 1976).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 146–151. From Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 13–18.

THEORETICAL AND MATHEMATICAL PHYSICS

On the Interaction of Point Charges in an Arbitrary Domain*

A. Silbergleit, I. Nemenman, and I. Mandel

Gravity Probe B, W.W. Hansen Experimental Physics Laboratory,
Stanford University, Stanford, CA 94305-4085, USA
e-mail: gleit@relgyro.stanford.edu, nemenman@itp.ucsb.edu, ilya@caltech.edu

Received August 29, 2002

Abstract—We develop a systematic approach to calculating the electrostatic force between point charges in an arbitrary domain with arbitrary boundary conditions. When the boundary is present, the simple expression for the force acting on a charge as "the charge times the field it is placed in" becomes ill-defined. However, this rule can be salvaged if the field in question is redefined to only include all terms that do not diverge at the charge position, in particular, those due to the charge itself. The proof requires handling the self-action energy divergence for point charges, which is accomplished by means of a geometrical regularization. © *2003 MAIK "Nauka/Interperiodica".*

I. INTRODUCTION

It is trivial to determine the force exerted by an external field[1] on a point charge in an otherwise empty space: by definition, "the force is equal to the charge times the field it is placed in." In particular, if the field in question is created by some other point charges, this rule, known by many from high school, still holds.

However, the situation changes drastically when a set of point charges creates the field inside an arbitrary domain with a boundary of some physical origin (reflected in the appropriate boundary conditions). Now the very notion of "the field the charge is placed in" becomes ill-defined. For example, a naive treatment of a single-charge problem might lead one to the entirely wrong conclusion that, since the entire field in the problem is due to the charge itself (there are no other sources!), "the field it is placed in" is zero, so there is no force at all.

A slightly more sophisticated physicist would argue that only that part of the field which diverges as $1/r^2$ near the charge is really created by it, while the rest is due to the boundary conditions, which represent mathematically the rearrangement of *other* physical charges at the boundary. Therefore, it is precisely what remains after subtracting the singular part that now gives "the field the charge is placed in." Unfortunately, such treatment leaves one in a somewhat awkward position of, first, calculating potentials and fields rigorously and, then, lowering the plank and using hand-waving arguments to derive forces from them. It is also not clear whether the conjecture about which part of the total field contributes to the force is always valid.

Thus, it seems appealing to show that the physical arguments can be backed by an accurate mathematical proof demonstrating that the *adjusted* rule, "the force is equal to the charge times the part of the field that does not diverge at the charge's location," is either universal or limited by certain conditions. To do this, one should turn to the most fundamental energy conservation argument which gives the force as the negative gradient of the energy in the charge's position. This approach also does not turn out to be straightforward, since the energy is infinite in the presence of point charges due to their self-action.

Perhaps because of these difficulties, as well as the misleading apparent simplicity of the problem, our literature search, which encompassed, in particular, [1–10] and many other books on the subject, revealed no ready result (except in [6], which we discuss in Section 4). So we give a careful derivation of the general expression for the force on point charges in this paper. It consists of the regularization of the problem, calculation of the force from the (regularized and finite) energy, and then taking the singular limit. The result agrees with one's intuitive expectations.

II. ELECTROSTATICS PROBLEM WITH VOLUME POINT CHARGES: POTENTIAL AND ENERGY

Consider an arbitrary three-dimensional domain D with a perfectly conducting boundary S and some N point electrical charges inside. The electrical potential $\psi(\mathbf{r})$, in this case, is determined by the following Dirichlet boundary value problem (we use SI units throughout the paper):

$$\Delta\psi = -\frac{1}{\varepsilon_0}\sum_{i=1}^{N} q_i\delta(\mathbf{r}-\mathbf{r}_i), \quad \mathbf{r}, \mathbf{r}_i \in D; \tag{1}$$

$$\psi|_S = 0. \tag{2}$$

[1] In fact, the word "external" means that the field is produced by some independent sources, and the boundaries, if any, are far away from the charge.

*This article was submitted by the authors in English.

1063-7842/03/4802-0146$24.00 © 2003 MAIK "Nauka/Interperiodica"

Here, $\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z$ is the vector radius of a point, and $\mathbf{r}_i = x_i\mathbf{e}_x + y_i\mathbf{e}_y + z_i\mathbf{e}_z$ specifies the ith charge position, with $\mathbf{e}_\alpha$, $\alpha = x, y, z$, being the unit vectors in the direction of the corresponding Cartesian axes.

By the superposition principle, the potential $\psi(\mathbf{r})$ is merely the sum of the potentials induced by each charge separately,

$$\psi(\mathbf{r}) = \kappa\sum_{j=1}^{N} q_j G(\mathbf{r}, \mathbf{r}_j), \tag{3}$$

$$\equiv \kappa\sum_{j=1}^{N} q_j\left[\frac{1}{|\mathbf{r}-\mathbf{r}_j|} + G_R(\mathbf{r}, \mathbf{r}_j)\right], \tag{4}$$

where $\kappa = 1/4\pi\varepsilon_0$, and $G_R(\mathbf{r}, \mathbf{r}_j)$ is the regular part of the Green's function $G(\mathbf{r}, \mathbf{r}_j)$ of the corresponding boundary value problem [set $q_j = 1$, $q_i = 0$, $i \neq j$ in Eq. (1)]. Both functions are, of course, symmetric in their arguments,

$$G(\mathbf{r}, \mathbf{r}_j) = G(\mathbf{r}_j, \mathbf{r}), \quad G_R(\mathbf{r}, \mathbf{r}_j) = G_R(\mathbf{r}_j, \mathbf{r}). \tag{5}$$

Furthermore, we can rewrite Eq. (4), splitting the potential in a sum of its singular and regular parts,

$$\psi(\mathbf{r}) = \kappa\sum_{j=1}^{N}\frac{q_j}{|\mathbf{r}-\mathbf{r}_j|} + \psi_R(\mathbf{r}), \tag{6}$$

$$\psi_R(\mathbf{r}) \equiv \kappa\sum_{j=1}^{N} q_j G_R(\mathbf{r}, \mathbf{r}_j), \tag{7}$$

where $\psi_R(\mathbf{r})$ is a regular function satisfying the Laplace equation everywhere in D [by continuity, this holds also at any regular point[2] of the boundary S, although this is irrelevant to our discussion]. Note that both the potential ψ and its regular part ψ_R actually depend on the positions of the charges $\mathbf{r}_i$, as well as on the observation point $\mathbf{r}$, which is reflected in the full notation

$$\psi(\mathbf{r}) \equiv \psi(\mathbf{r}, \mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_N), \tag{8}$$

$$\psi_R(\mathbf{r}) \equiv \psi_R(\mathbf{r}, \mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_N). \tag{9}$$

We assume that the potential is known, and we are interested in finding the force $\mathbf{F}^i$ acting on the charge q_i. From the energy conservation for the considered problem, the force is given by (cf. [6])

$$\mathbf{F}^i = -\frac{\partial}{\partial\mathbf{r}_i}W_D, \quad \frac{\partial}{\partial\mathbf{r}_i} = \frac{\partial}{\partial x_i}\mathbf{e}_x + \frac{\partial}{\partial y_i}\mathbf{e}_y + \frac{\partial}{\partial z_i}\mathbf{e}_z, \tag{10}$$

where W_D is the energy of the field in the volume D,

$$W_D = \frac{\varepsilon_0}{2}\int_D (\nabla\psi)^2 dV. \tag{11}$$

[2] We allow for boundary singularities, such as sharp edges and spikes, provided that the Meixner-type finite energy condition [11] is satisfied near them; in particular, the domain D can be infinite.

Note that we alternatively write ∇ or $\partial/\partial\mathbf{r}$ for the gradient, whatever seems proper in a particular expression.

The problem is, however, that the above integral obviously diverges due to self-interaction of the point charges (the energy of a single point charge is infinite). We are going to show that even though the energy for a given point charge distribution is infinite, the *difference* between its two values corresponding to any two charge configurations *is finite* (for an arbitrary boundary shape) and goes to zero when one charge distribution tends to the other. Hence, the *force is also finite*, in accordance with common intuition. The situation with energy here is similar to the one arising in the calculation of the Casimir effect [12], which also requires some generalization.

III. REGULARIZED ENERGY AND THE FORCE ON THE CHARGES

We surround each volume charge q_i by a small sphere S_i^ϵ of radius ϵ; we write D_i^ϵ for the ball inside it. We define D^ϵ as D without all domains D_i^ϵ, and S^ϵ as a union of S and all spherical surfaces S_i^ϵ (see Fig. 1). In effect, S^ϵ is the boundary of the domain D^ϵ and $D^\epsilon \longrightarrow D$, $S^\epsilon \longrightarrow S$ when $\epsilon \longrightarrow 0$.

Using Eq. (10), we may now define the force acting on the charge q_i as

$$\mathbf{F}^i = \lim_{\epsilon\to 0}\mathbf{F}_\epsilon^i = -\lim_{\epsilon\to 0}\frac{\partial}{\partial\mathbf{r}_i}W_D^\epsilon, \tag{12}$$

where W_D^ϵ is the regularized energy, that is, the energy of the field in D^ϵ, which is finite. It is important to note the order of operations in Eq. (12): first, take the gradient of the regularized energy in the charge position, then, take the (singular) limit. In principle, we also have to show that the final result does not depend on the regularization chosen, but this task is not easy. We will return to this problem briefly later in this paper.

In view of Eq. (11) and the fact that the total potential given by Eq. (3) or Eq. (6) is regular in D^ϵ, the reg-

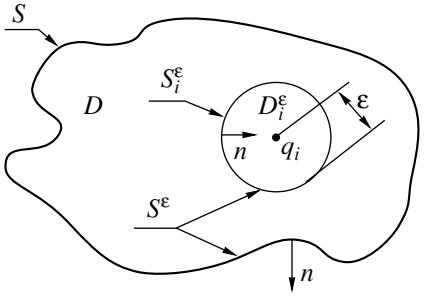


Fig. 1. Volumes, surfaces, and normal directions involved.

ularized energy is

$$W_D^\epsilon \equiv \frac{\varepsilon_0}{2}\int_{D^\epsilon}(\nabla\psi)^2 dV \tag{13}$$

$$= \frac{\varepsilon_0}{2}\int_{S^\epsilon}\psi\frac{\partial\psi}{\partial n}dA - \frac{\varepsilon_0}{2}\int_{D^\epsilon}\psi\Delta\psi dV, \tag{14}$$

where n is the direction of the *outward* normal to S^ϵ (and thus the *inward* normal to the spheres S_i^ϵ). For an infinite domain D, it is assumed here that the potential and its gradient drop at infinity fast enough to make the contribution of integrating over the sphere of a large radius vanishing in the limit, an assumption which has to be verified in each particular case.

Since ψ is harmonic everywhere in D^ϵ, the volume integral on the right of the previous equality vanishes; the remaining surface integral is represented as

$$W_D^\epsilon = \frac{\varepsilon_0}{2}\sum_{k=1}^{N}\int_{S_k^\epsilon}\psi\frac{\partial\psi}{\partial n}dA + \frac{\varepsilon_0}{2}\int_{S}\psi\frac{\partial\psi}{\partial n}dA \tag{15}$$

and then, because of the boundary condition, Eq. (2), as

$$W_D^\epsilon = \frac{\varepsilon_0}{2}\sum_{k=1}^{N}\int_{S_k^\epsilon}\psi\frac{\partial\psi}{\partial n}dA. \tag{16}$$

We are ultimately interested in the limit $\epsilon \longrightarrow 0$, so we need to calculate only the quantities which do not vanish in this limit. The area of integration in each term of the above sum is $O(\epsilon^2)$; therefore, we need to keep track of the integrands that grow at least quadratically in ϵ^{-1}. Bearing this in mind and using Eq. (6) for the potential, we can write the surface integral in Eq. (16) as

$$\int_{S_k^\epsilon}\psi\frac{\partial\psi}{\partial n}dA = \int_{S_k^\epsilon}\frac{\kappa q_k}{|\mathbf{r}-\mathbf{r}_k|}\frac{\partial}{\partial n}\left(\frac{\kappa q_k}{|\mathbf{r}-\mathbf{r}_k|}\right)dA + \int_{S_k^\epsilon}\left[\psi_R(\mathbf{r}) + \sum_{j=1,\,j\neq k}^{N}\frac{\kappa q_j}{|\mathbf{r}-\mathbf{r}_j|}\right]\frac{\partial}{\partial n}\left(\frac{\kappa q_k}{|\mathbf{r}-\mathbf{r}_k|}\right)dA + O(\epsilon). \tag{17}$$

The first term in the above expression is, in fact, the regularized self-energy of the kth charge, $W_{k,\,\mathrm{self}}^\epsilon$. Performing elementary integration, we immediately find that

$$W_{k,\,\mathrm{self}}^\epsilon = \frac{\varepsilon_0}{2}\kappa^2\frac{4\pi q_k^2}{\epsilon} = \frac{\kappa q_k^2}{2\epsilon}. \tag{18}$$

The only feature of the regularized self-energy given by Eq. (18) that is important for our derivation is that *it does not depend on the position of the charge q_k, i.e., on the vector radius $\mathbf{r}_k$.*

The second term of the right-hand side of Eq. (17) can also be simplified if one notices that both ψ_R and $1/|\mathbf{r}-\mathbf{r}_j|$, $j \neq k$, are regular on S_k^ϵ and in D_k^ϵ. Therefore, their change within the small surface S_k^ϵ is of order ϵ. Thus, Eq. (17) may be rewritten as

$$\frac{\varepsilon_0}{2}\int_{S_k^\epsilon}\psi\frac{\partial\psi}{\partial n}dA = W_{k,\,\mathrm{self}}^\epsilon + \frac{\varepsilon_0}{2}\left[\psi_R(\mathbf{r}_k) + \sum_{j=1,\,j\neq k}^{N}\frac{\kappa q_j}{|\mathbf{r}_k-\mathbf{r}_j|}\right] \times \int_{S_k^\epsilon}\frac{\partial}{\partial n}\left(\frac{\kappa q_k}{|\mathbf{r}-\mathbf{r}_k|}\right)dA + O(\epsilon) \tag{19}$$

$$= W_{k,\,\mathrm{self}}^\epsilon + \frac{q_k}{2}\left[\psi_R(\mathbf{r}_k) + \sum_{j=1,\,j\neq k}^{N}\frac{\kappa q_j}{|\mathbf{r}_k-\mathbf{r}_j|}\right] + O(\epsilon),$$

and the integration here yielding the factor 4π is again elementary. This asymptotic equality may be differentiated in $\mathbf{r}_i$ with the same estimate of the remaining term.

Introducing now the previous expression into Eq. (16), we obtain

$$W_D^\epsilon = \sum_{k=1}^{N}W_{k,\,\mathrm{self}}^\epsilon + \frac{\kappa}{2}\sum_{k=1}^{N}\sum_{j=1,\,j\neq k}^{N}\frac{q_j q_k}{|\mathbf{r}_j-\mathbf{r}_k|} + \frac{1}{2}\sum_{k=1}^{N}q_k\psi_R(\mathbf{r}_k) + O(\epsilon). \tag{20}$$

Equation (20), in its turn, is inserted in Eq. (12) for the force; as shown, the self-energies do not depend on the charge positions; hence, although diverging in the limit $\epsilon \longrightarrow 0$, *they do not contribute to the force.* The rest is pretty straightforward, except that one has to be careful when differentiating the last term on the right of Eq. (20) with $k = i$: as is seen from Eq. (8), in this case $\mathbf{r}_i$ stands for *two* (and not one!) arguments of ψ_R, namely, $\psi_R(\mathbf{r}_i) \equiv \psi_R(\mathbf{r}_i, \mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_N)$, and *both of them* have to be differentiated. Bearing this in mind, the expression for the force finally becomes

$$\mathbf{F}^i = -\kappa q_i\sum_{k=1,\,k\neq i}^{N}\frac{\partial}{\partial\mathbf{r}_i}\frac{q_k}{|\mathbf{r}_i-\mathbf{r}_k|} - \frac{1}{2}\left[\sum_{k=1}^{N}q_k\frac{\partial}{\partial\mathbf{r}_i}\psi_R(\mathbf{r})\big|_{\mathbf{r}=\mathbf{r}_k} + q_i\frac{\partial}{\partial\mathbf{r}}\psi_R(\mathbf{r})\big|_{\mathbf{r}=\mathbf{r}_i}\right] = \kappa q_i\sum_{k=1,\,k\neq i}^{N}q_k\frac{\mathbf{r}_i-\mathbf{r}_k}{|\mathbf{r}_i-\mathbf{r}_k|^3} \tag{21}$$

$$-\frac{1}{2}\left[\sum_{k=1}^{N} q_k \frac{\partial}{\partial \mathbf{r}_i}\psi_R(\mathbf{r})|_{\mathbf{r}=\mathbf{r}_k} + q_i \frac{\partial}{\partial \mathbf{r}}\psi_R(\mathbf{r})|_{\mathbf{r}=\mathbf{r}_i}\right].$$

This is the general result for the electrostatics which can be transformed further. Indeed, the direct substitution of the expression for ψ_R from Eq. (6) into Eq. (21) provides the force in the form

$$\mathbf{F}^i = -\kappa q_i \sum_{k=1,\,k\neq i}^{N} \frac{\partial}{\partial \mathbf{r}_i}\frac{q_k}{|\mathbf{r}_i - \mathbf{r}_k|}$$

$$-\frac{\kappa q_i}{2}\left[\sum_{k=1}^{N} q_k \frac{\partial}{\partial \mathbf{r}_i} G_R(\mathbf{r}, \mathbf{r}_i)|_{\mathbf{r}=\mathbf{r}_k} + \sum_{j=1}^{N} q_j \frac{\partial}{\partial \mathbf{r}} G_R(\mathbf{r}, \mathbf{r}_j)|_{\mathbf{r}=\mathbf{r}_i}\right] \quad (22)$$

$$= -\kappa q_i\left[\sum_{k=1,\,k\neq i}^{N} \frac{\partial}{\partial \mathbf{r}_i}\frac{q_k}{|\mathbf{r}_i - \mathbf{r}_k|} + \sum_{k=1}^{N} q_k \frac{\partial}{\partial \mathbf{r}} G_R(\mathbf{r}, \mathbf{r}_k)|_{\mathbf{r}=\mathbf{r}_i}\right],$$

and here we have used the symmetry property of Eq. (5) to obtain the second equality. To make the result even more physically transparent, we rewrite Eq. (22), in its turn, in the following way:

$$\mathbf{F}^i = -\kappa q_i \nabla\left\{\sum_{k=1}^{N} q_k\left[\frac{1}{|\mathbf{r}-\mathbf{r}_k|} + G_R(\mathbf{r}, \mathbf{r}_k)\right] - \frac{q_i}{|\mathbf{r}-\mathbf{r}_i|}\right\}\Bigg|_{\mathbf{r}=\mathbf{r}_i} = -q_i \nabla\left[\psi(\mathbf{r}) - \frac{\kappa q_i}{|\mathbf{r}-\mathbf{r}_i|}\right]\Bigg|_{\mathbf{r}=\mathbf{r}_i}. \quad (23)$$

Note that the last expression, indeed, coincides with our intuitive conjecture about the form of the force.

IV. DISCUSSION

Our first remark on the expressions for the force in Eqs. (21)–(23) is that for the charges in a free space (volume D is the whole space, no boundaries are present), apparently, $G_R(\mathbf{r}, \mathbf{r}_k) \equiv 0$, $\psi_R \equiv 0$, and the classical Coulomb formula for the force is restored.

Next, Eq. (23) shows that the rule "the force is the charge times the field it is placed in" does work *if one counts the regular part of the field produced by the charge in question as a part of the "field the charge is placed in."* It also allows for a certain "minimal principle"; namely, to get the right answer for the force, one should remove from the field only the part which otherwise makes the result infinite, *and nothing beyond that.* As we mentioned in the Introduction, this result is supported by physical intuition. It becomes even more obvious if one notes that the singular part of the field removed is radial, and a radial field produces no force.

The contribution of the regular part of the field created by a charge to the force acting on it is especially important in the case of a single charge, as one may see from the simplest example of a charge near a conducting plane. It is precisely the regular part of the field produced by the charge in question (equal to the field of the image charge) that gives the whole answer when no other charges are present. Finally, an important question is how *robust* our regularization of the problem is, i.e., whether the result for the force changes or not if one uses a different regularization. There are two significant points demonstrating such robustness.

The first one is concerned with the *geometrical* regularization that we used. If one chooses domain D_k^ϵ around q_k to be not a ball but some differently shaped volume bounded by a smooth surface S_k^ϵ ("topological ball"), then it is not difficult to see that all the terms in Eq. (20) for the regularized energy remain unchanged, and, hence, our result for the force is still valid. This can be demonstrated in exactly the same way as above, only the computation of the integral over the surface S_k^ϵ in Eq. (19) requires a well-known result from potential theory (cf. [13], n. 193, or [14]).

As for the first integral on the right of Eq. (17), which defines the self-energy $W_{k,\,\mathrm{self}}^\epsilon$, its explicit expression is not required, and its only relevant property, namely, its independence of $\mathbf{r}_k$, is obvious.

An alternative way of regularization, so widely used during the whole "pre-Dirac delta-function" era, is *physical* regularization, where the point charge q_k is replaced, within a small volume D_k^ϵ, with some smooth charge distribution of the density $\rho_k^\epsilon(\mathbf{k})$ and the same total charge q_k, and ϵ is taken to be zero in the answer. From a technical point of view, this approach proves to be more complicated in this particular case, but it leads again to the same terms in Eq. (20) for the regularized energy. The key point here is to start with the following expression for the regularized energy,

$$W_D^\epsilon \equiv \frac{1}{2}\int_D \rho^\epsilon \psi\, dV = \frac{1}{2}\sum_{k=1}^{N}\int_{D_k^\epsilon} \rho_k^\epsilon \psi\, dV, \quad (24)$$

and then, instead of Eq. (3), split the potential into a sum of volume potentials of $\rho_k^\epsilon(\mathbf{r})$ over D_k^ϵ (which becomes singular in the limit) and a regular addition $\psi_R^\epsilon(\mathbf{r})$.

In particular, this regularization is used by Smythe in Section 3.08 of [6] for calculating the force on a single point charge in a domain with zero potential at the boundary. In that work, derivation is at the "physical level of accuracy" and the answer is not brought down to its physically most relevant form of Eq. (23). Moreover, the final answer [right-hand side of Eq. (2) in that section] is, unfortunately, formally diverging because of the inappropriate use of the notation for the total potential in the place where its regular part should be.

Finally, we want to end our discussion by mentioning that the electrostatic problem we just solved, as well as its generalizations (see Section 5), involve only volume charges. On the other hand, magnetostatic problems that deal, for example, with magnetic fluxes trapped in superconducting media (cf. [15]) give rise to surface charges. Analysis of these is of extreme importance for modern experimental physics [16]. No easy solution for the force between surface charges should be anticipated since the details of the boundary shape, such as its curvature, are expected to play a role; the interaction of such surface charges will be discussed elsewhere.

V. GENERALIZATION: OTHER BOUNDARY CONDITIONS

We can now generalize our result for other conditions at the boundary.

A modest but potentially useful generalization may be applied to the case of electrodes, where an arbitrary distribution of the potential $V(\mathbf{r})$, and not just a zero, is specified at the boundary:

$$\psi|_S = V(\mathbf{r}), \quad \mathbf{r} \in S. \tag{25}$$

Let us split the potential in two,

$$\psi = \psi^{(1)} + \psi^{(2)}, \tag{26}$$

in which the first is caused by point charges without any voltage applied to the boundary, and the second is entirely due to the boundary voltage. Therefore, $\psi^{(1)}$ satisfies the boundary value problem of Eqs. (1) and (2),

$$\Delta\psi^{(1)} = -\frac{1}{\varepsilon_0}\sum_{i=1}^{N} q_i\delta(\mathbf{r}-\mathbf{r}_i), \quad \mathbf{r}, \mathbf{r}_i \in D; \tag{27}$$

$$\psi^{(1)}|_S = 0. \tag{28}$$

According to what was proven above, the force on a charge from the field specified by the potential $\psi^{(1)}$ is given according to Eq. (23),

$$\mathbf{F}^i_{(1)} = -q_i\nabla\left[\psi^{(1)}(\mathbf{r}) - \frac{\kappa q_i}{|\mathbf{r}-\mathbf{r}_i|}\right]\Bigg|_{\mathbf{r}=\mathbf{r}_i}. \tag{29}$$

On the other hand, the potential $\psi^{(2)}$, satisfying

$$\Delta\psi^{(2)} = 0, \quad \mathbf{r} \in D; \quad \psi^{(2)}|_S = V(\mathbf{r}), \tag{30}$$

describes a field *external* to all point charges, since it does not depend on them and their positions. Therefore, the force exerted by this field is

$$\mathbf{F}^i_{(2)} = -q_i\nabla\psi^{(2)}(\mathbf{r})|_{\mathbf{r}=\mathbf{r}_i}. \tag{31}$$

Using the superposition principle, we add these two forces to reinstate the result of Eq. (23) in the considered case:

$$\mathbf{F}^i = \mathbf{F}^i_{(1)} + \mathbf{F}^i_{(2)} = -q_i\nabla\left[\psi(\mathbf{r}) - \frac{\kappa q_i}{|\mathbf{r}-\mathbf{r}_i|}\right]\Bigg|_{\mathbf{r}=\mathbf{r}_i}. \tag{32}$$

The mixed boundary conditions

$$\psi|_{S_1} = V(\mathbf{r}), \quad \varepsilon_0\frac{\partial\psi}{\partial n}\Big|_{S_2} = \sigma(\mathbf{r}), \tag{33}$$

where the surfaces S_1, S_2 are nonintersecting ($S_1 \cap S_2 = \varnothing$) and comprise the whole boundary ($S_1 \cup S_2 = S$), and $V(\mathbf{r})$, $\sigma(\mathbf{r})$ are given functions, lead to the same standard result for the force [Eq. (23)] without any new technical difficulties. Indeed, we split the total potential in two as in Eq. (26) and require that

$$\Delta\psi^{(1)} = -\frac{1}{\varepsilon_0}\sum_{i=1}^{N} q_i\delta(\mathbf{r}-\mathbf{r}_i), \quad \mathbf{r}, \mathbf{r}_i \in D; \tag{34}$$

$$\psi^{(1)}|_{S_1} = 0, \quad \frac{\partial\psi^{(1)}}{\partial n}\Big|_{S_2} = 0 \tag{35}$$

and

$$\Delta\psi^{(2)} = 0, \quad \mathbf{r} \in D; \tag{36}$$

$$\psi^{(2)}|_{S_1} = V(\mathbf{r}), \quad \varepsilon_0\frac{\partial\psi^{(2)}}{\partial n}\Big|_{S_2} = \sigma(\mathbf{r}). \tag{37}$$

The derivation of the force from $\psi^{(1)}$ is performed exactly as in Section 3 and leads to Eq. (29). The field that is external to the charges from $\psi^{(2)}$ produces the force of Eq. (31), so by superposition the total force is again the same as in Eq. (23) [or Eq. (32)].

The appropriate splitting of the potential into two parts [Eq. (26)] is somewhat more difficult for the Neumann boundary condition,

$$\varepsilon_0\frac{\partial\psi}{\partial n}\Big|_S = \sigma(\mathbf{r}), \int_S \sigma(\mathbf{r})dA + Q = 0, \quad Q \equiv \sum_{j=1}^{N} q_j; \tag{38}$$

namely, the solvability criterion (the total charge must be zero) requires that, when splitting the potential, another charge Q (equal to the sum of the point charges q_i) be added and subtracted at some point $\mathbf{r}_*$ of the domain D to obtain

$$\Delta\psi^{(1)} = -\frac{1}{\varepsilon_0}\left[\sum_{i=1}^{N} q_i\delta(\mathbf{r}-\mathbf{r}_i) - Q\delta(\mathbf{r}-\mathbf{r}_*)\right], \quad \mathbf{r}, \mathbf{r}_i, \mathbf{r}_* \in D; \tag{39}$$

$$\frac{\partial\psi^{(1)}}{\partial n}\Big|_S = 0, \tag{40}$$

as well as

$$\Delta\psi^{(2)} = -\frac{Q}{\varepsilon_0}\delta(\mathbf{r}-\mathbf{r}_*), \quad \mathbf{r}, \mathbf{r}_* \in D; \tag{41}$$

$$\varepsilon_0 \frac{\partial\psi^{(2)}}{\partial n}\bigg|_S = \sigma(\mathbf{r}) \tag{42}$$

with both problems solvable. Again, the derivation of the force from $\psi^{(1)}$ satisfying the homogeneous boundary condition is performed exactly as before and leads to Eq. (29), the field $\psi^{(2)}$ external to the charges exerts the force given in Eq. (31), and the result of Eq. (23) holds by superposition. The problem itself, though, is not too realistic, except for the case of an insulated boundary, $\sigma(\boldsymbol{r}) \equiv 0$.

ACKNOWLEDGMENTS

This work was supported by NASA, grant no. NAS 8-39225 to Gravity Probe B. In addition, I.N. was partially supported by the NEC Research Institute. The authors are grateful to R.V. Wagoner and V.S. Mandel for valuable remarks, and to the GP-B Theory Group for fruitful discussions.

REFERENCES

1. A. J. W. Sommerfeld, *Electrodynamics* (Academic Press, New York, 1952; Inostrannaya Literatura, 1958).
2. I. E. Tamm, *The Principles of Electricity Theory* (GITTL, Moscow, 1957).
3. J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941; Gostekhizdat, Moscow, 1948).
4. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed. (GIFML, Moscow, 1962; Pergamon, Oxford, 1975).
5. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Pergamon, New York, 1984).
6. W. R. Smythe, *Static and Dynamic Electricity* (Hemisphere, New York, 1989).
7. J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1999).
8. R. P. Feynman, *The Feynman Lectures on Physics*, Vol. 2: *Electromagnetism and Matter* (Addison-Wesley, Redwood City, 1989).
9. J. M. Crowley, *Fundamentals of Applied Electrostatics* (Wiley, New York, 1986).
10. *Electrostatics and Its Applications*, Ed. by A. D. Moore (Wiley, New York, 1973).
11. R. Mittra and S. W. Lee, *Analytical Techniques in the Theory of Guided Waves* (Springer, Berlin, 1967).
12. V. M. Mostepanenko and N. N. Trunov, *The Casimir Effect and Its Applications* (Clarendon, Oxford, 1997).
13. V. I. Smirnov, *A Course of Higher Mathematics* (GITTL, Moscow, 1957; Addison-Wesley, Reading, Mass., 1964), Vol. 4.
14. O. D. Kellogg, *Foundations of Potential Theory* (Macmillan, New York, 1971).
15. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1996).
16. I. M. Nemenman and A. S. Silbergleit, J. Appl. Phys. **86**, 614 (1999).

Technical Physics, Vol. 48, No. 2, 2003, pp. 152–164. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 19–30.

GASES AND LIQUIDS

On Internal Mode Resonance in a Nonlinearly Vibrating Volumetrically Charged Dielectric Drop

S. O. Shiryaeva

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

e-mail: shir@uniyar.ac.ru

Received July 18, 2002

Abstract—In the approximation quadratic in the amplitude of an arbitrary initial deformation of an equilibrium spherical uniformly (volumetrically) charged drop of a dielectric liquid, an analytical expression for the drop surface generatrix as a function of time is derived in the case when the drop shape executes axisymmetric vibrations. A condition that must be imposed on mode frequencies in order for resonant interaction between modes to take place in the quadratic approximation is found. It is shown that many resonances, rather than one known previously, are realized when the self-charge is insufficient (subcritical) for drop surface instability against self-charge to arise. Nonlinear two- and three-mode resonant interactions are studied. *© 2003 MAIK "Nauka/Interperiodica".*

(1) The nonlinear capillary vibration of charged drops is of considerable interest in a variety of areas of science and technology (see, e.g., [1–11] and Refs. therein). However, because of the cumbersome analytical calculations, many aspects of the nonlinear vibration of charged drops are as yet little understood. The issues that remain poorly studied are, for example, the so-called translational instability of drops, which shows up when several neighboring modes arise in the mode spectrum excited at the initial time instant [11]; internal nonlinear resonant interaction between different modes of the capillary vibration of the drop [1, 8]; and the effect of the charge distribution over the drop volume on nonlinear vibration. In this work, we solve the problem of a nonlinearly vibrating uniformly (volumetrically) charged drop of a dielectric liquid, separate out resonances arising in the case of a subcritical (in terms of linear instability) charge, and analyze the solution in the vicinity of typical resonances.

(2) Consider the time evolution of the shape of a spherical drop of an ideal incompressible dielectric liquid with a density ρ, surface tension coefficient ζ, and permittivity ε_d. We assume that the drop of radius R is in a vacuum and that its total charge Q is uniformly distributed over the volume with a constant charge density μ_q. At the time instant $t = 0$, the equilibrium spherical shape of the drop undergoes an axisymmetric perturbation of fixed amplitude, the perturbation being much smaller than the radius of the drop. Our aim is to find the spectrum of arising capillary oscillations (i.e., the shape of the drop) for $t > 0$.

Let us assume that the drop is axisymmetric at any time instant. Then, in the spherical coordinate system with the origin at the center of mass of the drop, the equation describing the surface of the drop in dimensionless variables where $R = \rho = \zeta = 1$ has the form

$$r(\Theta, t) = 1 + \xi(\Theta, t), \quad |\xi| \ll 1. \tag{1}$$

The flow inside the drop is assumed to be potential. In this case, the velocity field in the drop $\mathbf{V}(\mathbf{r}, t) = \nabla\psi(\mathbf{r}, t)$ is completely defined by the velocity potential function $\psi(\mathbf{r}, t)$.

The set of equations for drop shape evolution consists of the Laplace equations for the velocity potential $\psi(\mathbf{r}, t)$ and external electrostatic potential $\Phi_{ex}(\mathbf{r}, t)$, as well as the Poisson equation for the internal electrostatic potential $\Phi_{in}(\mathbf{r}, t)$:

$$\Delta\psi(\mathbf{r}, t) = 0, \tag{2}$$

$$\Delta\Phi_{ex}(\mathbf{r}, t) = 0, \quad \Delta\Phi_{in}(\mathbf{r}, t) = -4\pi\frac{\mu_q}{\varepsilon_d}, \tag{3}$$

where Δ is Laplacian.

The boundary conditions are as follows:

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

$$\begin{gathered} r \longrightarrow \infty\colon \nabla\Phi_{ex}(\mathbf{r}, t) \longrightarrow 0; \\ r \longrightarrow 0\colon \nabla\Phi_{in}(\mathbf{r}, t) < \infty; \end{gathered} \tag{5}$$

$$r = 1 + \xi(\Theta, t)\colon \frac{\partial\xi}{\partial t} = \frac{\partial\psi}{\partial r} - \frac{1}{r^2}\frac{\partial\xi}{\partial\Theta}\frac{\partial\psi}{\partial\Theta}; \tag{6}$$

$$\begin{gathered} \Delta p - \frac{\partial\psi}{\partial t} - \frac{1}{2}(\nabla\psi)^2 - \mu_q\Phi_{in} + \frac{1}{8\pi}(\varepsilon_d - 1)(\nabla\Phi_{ex})^2 \\ - \frac{(\varepsilon_d - 1)^2}{8\pi\varepsilon_d}(\mathbf{n}\cdot\nabla\Phi_{ex})^2 = \mathrm{div}\mathbf{n}; \end{gathered} \tag{7}$$

1063-7842/03/4802-0152$24.00

$$\Phi_{ex} = \Phi_{in}; \tag{8}$$

$$(\mathbf{n} \cdot \nabla\Phi_{ex}) = \varepsilon_d(\mathbf{n} \cdot \nabla\Phi_{in}). \tag{9}$$

We also introduce the obvious conditions of center-of-mass immobility and conservation of drop volume:

$$\int_V r^2 dr \sin\Theta d\Theta d\phi = \frac{4}{3}\pi, \quad V = [0 \le r \le 1 + \xi(\Theta, t),\ 0 \le \Theta \le \pi,\ 0 \le \phi \le 2\pi]; \tag{10}$$

$$\int_V \mathbf{e}_r \cdot r^3 dr \sin\Theta d\Theta d\phi = 0, \quad V = [0 \le r \le 1 + \xi(\Theta, t),\ 0 \le \Theta \le \pi,\ 0 \le \phi \le 2\pi]. \tag{11}$$

Note that conditions (10) and (11) have to be satisfied at any time instant, including the initial one. Therefore, at $t = 0$, they define the amplitudes of the zero and first modes in the expansion of the initial perturbation $\xi(\Theta)$ of the equilibrium spherical shape of the drop in Legendre polynomials. In other words, the amplitudes of the zero and first modes cannot be arbitrary: they depend on the initial deformation.

The initial conditions are set as the initial deformations of the equilibrium spherical shape of the drop and as the zero initial velocity of the surface:

$$t = 0: \xi(\Theta) = \xi_0 P_0(\mu) + \xi_1 P_1(\mu) + \varepsilon \sum_{i \in \Xi} h_i P_i(\mu); \quad \frac{\partial\xi(\Theta)}{\partial t} = 0; \quad \sum_{i \in \Xi} h_i = 1, \tag{12}$$

where Ξ is a set of the numbers of initially excited vibrational modes and $\mu \equiv \cos\Theta$.

In (6)–(12), Δp is the difference in the constant pressures inside and outside of the drop in equilibrium; $\mathbf{n}$, the unit normal vector to the surface defined by (1); ε, the amplitude of the initial perturbation of the surface shape (small parameter of the problem); $P_i(\mu)$, ith-order Legendre polynomials; h_i, coefficients specifying the partial contribution from an ith mode to the total initial perturbation; and ξ_0 and ξ_1, constants that are determined from conditions (10) and (11) at the initial time instant and are given by

$$\xi_0 \approx -\varepsilon^2 \sum_{i \in \Xi} \frac{h_i^2}{(2i+1)} + O(\varepsilon^3); \quad \xi_1 \approx -\varepsilon^2 \sum_{i \in \Xi} \frac{9ih_{i-1}h_i}{(2i-1)(2i+1)} + O(\varepsilon^3) \tag{13}$$

up to second-order infinitesimals in ε.

In Eqs. (3), (5), (8), and (9), the electrical charge is assumed to be "frozen" in the liquid, i.e., uniformly distributed over the drop volume with a constant volumetric density $\mu_q = (3Q/4\pi)$.

(3) To solve the problem posed, we will use the classical method of multiple scales. The desired functions $\xi(\Theta, t)$, $\psi(\mathbf{r}, t)$, $\Phi_{ex}(\mathbf{r}, t)$, and $\Phi_{in}(\mathbf{r}, t)$ are expanded in the power series in the small parameter ε that are considered as functions of various time scales (rather than merely of time t) defined via the small parameter ε: $T_m \equiv \varepsilon^m t$. Then,

$$\xi(\Theta, t) = \sum_{m=1}^{\infty} \varepsilon^m \xi^{(m)}(\Theta, T_0, T_1);$$

$$\psi(\mathbf{r}, t) = \sum_{m=1}^{\infty} \varepsilon^m \psi^{(m)}(r, \Theta, T_0, T_1);$$

$$\Phi_{ex}(\mathbf{r}, t) = \sum_{m=1}^{\infty} \varepsilon^m \Phi_{ex}^{(m)}(r, \Theta, T_0, T_1); \tag{14}$$

$$\Phi_{in}(\mathbf{r}, t) = \sum_{m=1}^{\infty} \varepsilon^m \Phi_{in}^{(m)}(r, \Theta, T_0, T_1).$$

The hydrodynamic part of the problem is solved in a similar way as the problem [9, 10] of the surface vibration of a charged drop of a perfectly conducting liquid. The only difference is requirement (11) that the center of mass be stationary at any time instant, which eventually defines the first mode amplitude (see (13)) in expansion (12) and the time evolution of the first mode (see (19) and (32)). The electrical part of the problem is considered at length in Appendix A.

We will restrict our analysis by the second-order approximation to find the desired parameters as a function of two time scales T_0 and T_1.

Substituting expansions (14) into set (2)–(11) and equating the terms with the same powers of ε, we arrive at a set of boundary-value problems for $\xi^{(m)}$, $\Psi^{(m)}$, $\Phi_{ex}^{(m)}$, and $\Phi_{in}^{(m)}$.

It is obvious that each of the functions $\Psi^{(m)}$, $\Phi_{ex}^{(m)}$, and $\Phi_{in}^{(m)}$ must satisfy linear equations (2) and (3). Note only that owing to the model of frozen charge, the potential $\Phi_{in}^{(0)}$ is the solution to Poisson equation (3), while $\Phi_{in}^{(1)}$ and $\Phi_{in}^{(2)}$ are found from the homogeneous Laplace equation.

In the zeroth approximation, we will come to a solution for an equilibrium spherical dielectric drop with the easily calculable electric potentials of zeroth order of smallness $\Phi_{in}^{(0)}$ and $\Phi_{ex}^{(0)}$ (see Appendix A).

(4) Solutions to Eqs. (2) and (3) for the functions of first and second orders of smallness that satisfy bound-

ary conditions (4) and (5) can be written in the form

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

$$\Phi_{\text{ex}}^{(m)}(r, \Theta, T_0, T_1) = \sum_{n=0}^{\infty} F_n^{(m, \text{ex})}(T_0, T_1) r^{-(n+1)} P_n(\mu); \quad (15)$$

$$\Phi_{\text{in}}^{(m)}(r, \Theta, T_0, T_1) = \sum_{n=0}^{\infty} F_n^{(m, \text{in})}(T_0, T_1) r^n P_n(\mu).$$

Successive corrections to the equilibrium surface of the drop are also represented as expansions in Legendre polynomials:

$$xi^{(m)}(\Theta, T_0, T_1) = \sum_{n=0}^{\infty} M_n^{(m)}(T_0, T_1) P_n(\mu); \quad m = 1, 2. \quad (16)$$

Substituting solutions (15) and (16) at $m = 1$ into a set of first-order boundary conditions that is derived from (6)–(9), we come, after appropriate rearrangements, to differential equations for the coefficients $M_n^{(1)}(T_0, T_1)$:

$$\frac{\partial^2 M_n^{(1)}(T_0, T_1)}{\partial T_0^2} + \omega_n^2 M_n^{(1)}(T_0, T_1) = 0;$$

$$\omega_n^2 \equiv n(n-1)[(n+2) - W\kappa_1(n, \varepsilon_d)]; \quad W = \frac{Q^2}{4\pi}; \quad (17)$$

$$\kappa_1(n, \varepsilon_d) \equiv \frac{n\varepsilon_d^2 - (2n-5)\varepsilon_d + (n+1)}{\varepsilon_d(n\varepsilon_d + n + 1)}.$$

A solution to Eqs. (17) is harmonic functions with coefficients dependent on the time T_1:

$$M_n^{(1)}(T_0, T_1) = A_n^{(1)}(T_1)\exp(i\omega_n T_0) + \text{c.c.}; \quad n \geq 2;$$
$$A_n^{(1)}(T_1) = a_n^{(1)}(T_1)\exp(ib_n^{(1)}(T_1)) \quad (18)$$

(hereafter, "c.c." means terms complex conjugate to written terms), where $a_n^{(1)}(T_1)$ and $b_n^{(1)}(T_1)$ are real functions whose dependence on T_1 can be found only after the problem of the next order of smallness is considered.

From conditions (10) and (11) written in the approximation linear in ε, it follows that

$$M_0^{(1)}(T_0, T_1) = 0; \quad M_1^{(1)}(T_0, T_1) = 0. \quad (19)$$

Note that expressions (19) formally do not contradict Eqs. (17) for $n = 0$ and 1.

Satisfying boundary conditions (12) in the first approximation in ε, we obtain

$$a_i^{(1)}(0) = \frac{1}{2}h_i; \quad b_i^{(1)}(0) = 0; \quad i \in \Xi;$$
$$a_n^{(1)}(0) = 0; \quad b_n^{(1)}(0) \text{ is arbitrary}; \quad n \notin \Xi. \quad (20)$$

Substituting first-order solutions (18) and (19), as well as solutions (15) and (16) with $m = 2$, into a set of second-order boundary conditions that is derived from (6)–(9) and performing much more tedious rearrangements, we arrive at an equation for the coefficients $M_n^{(2)}(T_0, T_1)$:

$$\frac{\partial^2 M_n^{(2)}(T_0, T_1)}{\partial T^2} + \omega_n^2 M_n^{(2)}(T_0, T_1) = -2i\omega_n \frac{dA_n^{(1)}(T_1)}{dT_1}$$
$$\times \exp(i\omega_n T_0) + \sum_{l=2}^{\infty}\sum_{m=2}^{\infty} \{[\gamma_{lmn} + \omega_l\omega_m\eta_{lmn}]A_l^{(1)}(T_1)A_m^{(1)}(T_1)$$
$$\times \exp[i(\omega_l + \omega_m)T_0] + [\gamma_{lmn} - \omega_l\omega_m\eta_{lmn}]$$
$$\times A_l^{(1)}(T_1)\overline{A_m^{(1)}(T_1)}\exp[i(\omega_l - \omega_m)T_0]\} + \text{c.c.} \quad (21)$$

Hereafter, the bar means complex conjugation. The expressions for γ_{lmn} and η_{lmn} are given in Appendix B.

According to the general procedure [12], a uniformly suitable solution to Eq. (21) is found if terms responsible for secular terms and small-denominator terms in corresponding solutions are eliminated from the right of (21) by appropriately choosing the functions $A_n^{(1)}(T_1)$. The condition for eliminating these terms makes it possible to determine the form of the functions $A_n^{(1)}(T_1)$, i.e., the dependence of the first-order solution $M_n^{(1)}(T_0, T_1)$ on the "slow" time T_1.

Recall that in a partial solution to inhomogeneous equation (21), the terms in the inhomogeneous function that vary in the time scale T_0 with a frequency equal to the eigenfrequency ω_n of a solution to the corresponding homogeneous equation cause defects like secular and small-denominator terms.

(5) From the right of (21), it follows that if one of the relationships

$$\omega_p + \omega_q = \omega_k; \quad \omega_p - \omega_q = \omega_k \quad (22)$$

is valid for three vibrational modes with numbers p, q, and k, these modes are resonantly coupled. In this case, so-called secondary Raman resonance takes place.

It should be noted that, according to (17), the eigenfrequencies ω_n of drop vibration depend on the permittivity ε_d and drop charge (parameter W). At the critical value of W, $W_{\text{cr}} = 4/\kappa_1(2, \varepsilon_d)$, the vibration frequency of the fundamental mode (with $n = 2$) vanishes, indicating that the drop loses stability. This means that secondary

resonances may appear only if relationships (22) remain true at $W < W_{cr}$. In [1], one such resonance was found in the case $\omega_6 = 2\omega_4$.

Numerical calculations by formula (22) show that a greater number of secondary resonances may be excited in the drop. The values of W for which the relationship $\omega_3 = \omega_p - \omega_q$ is met at different p and q smaller than 10 with $\varepsilon_d = 5$ and $W_{cr} = 4.483$ are listed in Table 1 and with $\varepsilon_d = 1000$ and $W_{cr} = 4.004$, in Table 2. Similar data for the relationship $\omega_4 = \omega_p - \omega_q$ are listed in Tables 3 and 4. Note that in the formal limit $\varepsilon_d \longrightarrow \infty$, all the solutions obtained are reduced to solutions of a similar problem for a perfectly conducting drop with a total surface charge Q. Therefore, the data in Tables 2 and 4 for $\varepsilon_d = 1000$ should be considered to be valid for a conducting drop.

In the subsequent analysis of Eq. (21), we will consider the situation where modes p, q, and k obey the difference relationship $\omega_p - \omega_q = \omega_k$ and $p > q$. Clearly, the resonant sum is obtained from this difference relationship by transferring the frequency ω_q to the right-hand side of the latter and permutating p and k.

Let us consider separately four possible cases: $n \neq k$, p, q; $n = k$; $n = p$; and $n = q$. For each of them, the condition for eliminating secular and small-denominator terms from a solution to (21) is different.

For $n \neq k, p, q$, when the mode n is resonantly uncoupled, this condition has the simplest form:

$$\frac{dA_n^{(1)}(T_1)}{dT_1} = 0.$$

Using expression (18), where $A_n^{(1)}(T_1)$ is expressed through the scalar functions $a_n^{(1)}(T_1)$ and $b_n^{(1)}(T_1)$, and requiring the real and imaginary parts to vanish, one easily finds from the above equality that

$$\frac{da_n^{(1)}(T_1)}{dT_1} = \frac{db_n^{(1)}(T_1)}{dT_1} = 0.$$

This means that $a_n^{(1)}$ and $b_n^{(1)}$ do not depend on the slow time T_1 and therefore can be considered constant and equal to their initial values (20) in the second-order approximation. Expression (18) for the small-order coefficients $M_n^{(1)}(t)$ in the expansion of the equilibrium surface disturbance $\xi^{(1)}(\Theta, t)$ in Legendre polynomials takes the form

$$M_n^{(1)}(t) = \delta_{n,i} h_i \cos(\omega_i t); \quad i \in \Xi; \quad n \neq k, p, q, \quad (23)$$

where $\delta_{n,i}$ is the Kronecker symbol.

When analyzing Eq. (21) for modes with $n = k$, p, or q, we introduce the detuning parameter σ $O(1)$ defined as

$$\omega_p - \omega_q = \omega_k + \varepsilon\sigma$$

in order to underline the closeness of $\omega_p - \omega_q$ to ω_k.

Table 1. Mode numbers (q, $p \le 10$) excited at the initial time instant and the values of the parameter W related to the drop charge for which the resonant buildup of the third mode takes place ($\varepsilon_d = 5$, $W_{cr} = 4.483$)

$\varepsilon_d = 5$, $W_{cr} = 4.483$		
q	p	Q
4	5	4.355
5	6	3.867
6	7	3.361
7	8	2.847
8	9	2.328
9	10	1.807

Table 2. The same as in Table 1 for $\varepsilon_d = 1000$ and $W_{cr} = 4.004$

$\varepsilon_d = 1000$, $W_{cr} = 4.004$		
q	p	W
4	5	3.623
5	6	3.210
6	7	2.785
7	8	2.356
8	9	1.925
9	10	1.493

Table 3. The same as in Table 1 for the buildup of the fourth mode

$\varepsilon_d = 5$, $W_{cr} = 4.483$		
q	p	Q
3	5	4.355
4	6	3.248
5	7	2.018
6	8	0.750

Table 4. The same as in Table 2 for the buildup of the fourth mode

$\varepsilon_d = 1000$, $W_{cr} = 4.004$		
q	p	W
3	5	3.623
4	6	2.671
5	7	1.651
6	8	0.612

Then, the right of (21) will have terms containing the factors

$$\exp[i(\omega_p - \omega_q)T_0]$$
$$= \exp[i(\omega_k + \varepsilon\sigma)T_0] = \exp[i\sigma T_1]\exp[i\omega_k T_0];$$

$$\exp[i(\omega_k + \omega_q)T_0]$$
$$= \exp[i(\omega_p - \varepsilon\sigma)T_0] = \exp[-i\sigma T_1]\exp[i\omega_p T_0];$$
$$\exp[i(\omega_p - \omega_k)T_0]$$
$$= \exp[i(\omega_q + \varepsilon\sigma)T_0] = \exp[i\sigma T_1]\exp[i\omega_q T_0]$$

for the respective cases, and the conditions for eliminating secular terms from solutions to (21) for $n = k$, p, or q, respectively, will take the form

$$-2i\omega_k\frac{dA_k^{(1)}(T_1)}{dT_1} + \Lambda_{pqk}^{(-)}\exp[i\sigma T_1]A_p^{(1)}(T_1)\overline{A_q^{(1)}(T_1)} = 0;$$

$$-2i\omega_p\frac{dA_p^{(1)}(T_1)}{dT_1} + \Lambda_{kqp}^{(+)}\exp[-i\sigma T_1]A_k^{(1)}(T_1)A_q^{(1)}(T_1) = 0;$$
$$-2i\omega_q\frac{dA_q^{(1)}(T_1)}{dT_1} + \Lambda_{pkq}^{(-)}\exp[i\sigma T_1]A_p^{(1)}(T_1)\overline{A_k^{(1)}(T_1)} = 0; \quad (24)$$

$$\Lambda_{pqk}^{(\pm)} = (\gamma_{qpk} + \gamma_{qpk}) \pm \omega_p\omega_q(\eta_{pqk} + \eta_{qpk}).$$

Equating the real and imaginary parts of (24) to zero and introducing the new function

$$\beta_k^{(1)}(T_1) = \sigma T_1 - b_k^{(1)}(T_1), \quad (25)$$

we come to a set of differential equations for the real functions $a_k^{(1)}(T_1)$, $\beta_k^{(1)}(T_1)$, $a_p^{(1)}(T_1)$, $b_p^{(1)}(T_1)$, $a_q^{(1)}(T_1)$, and $b_q^{(1)}(T_1)$:

$$2\omega_k\frac{da_k^{(1)}(T_1)}{dT_1} = \Lambda_{pqk}^{(-)}a_p^{(1)}(T_1)a_q^{(1)}(T_1)\sin[\varphi_{kpq}^{(1)}(T_1)];$$

$$2\omega_k a_k^{(1)}(T_1)\frac{d\beta_k^{(1)}(T_1)}{dT_1} = 2\omega_k a_k^{(1)}(T_1)\sigma$$

$$+ \Lambda_{pqk}^{(-)}a_p^{(1)}(T_1)a_q^{(1)}(T_1)\cos[\varphi_{kpq}^{(1)}(T_1)];$$

$$2\omega_p\frac{da_p^{(1)}(T_1)}{dT_1} = -\Lambda_{kqp}^{(+)}a_k^{(1)}(T_1)a_q^{(1)}(T_1)\sin[\varphi_{kpq}^{(1)}(T_1)];$$

$$2\omega_p a_p^{(1)}(T_1)\frac{db_p^{(1)}(T_1)}{dT_1}$$
$$= -\Lambda_{kqp}^{(+)}a_k^{(1)}(T_1)a_q^{(1)}(T_1)\cos[\varphi_{kpq}^{(1)}(T_1)]; \quad (26)$$

$$2\omega_q\frac{da_q^{(1)}(T_1)}{dT_1} = \Lambda_{pkq}^{(-)}a_p^{(1)}(T_1)a_k^{(1)}(T_1)\sin[\varphi_{kpq}^{(1)}(T_1)];$$

$$2\omega_q a_q^{(1)}(T_1)\frac{db_q^{(1)}(T_1)}{dT_1}$$

$$= -\Lambda_{pkq}^{(-)}a_p^{(1)}(T_1)a_k^{(1)}(T_1)\cos[\varphi_{kpq}^{(1)}(T_1)];$$

$$\varphi_{kpq}^{(1)}(T_1) = \beta_k^{(1)}(T_1) + b_p^{(1)}(T_1) - b_q^{(1)}(T_1).$$

Relationships (20) serve as initial conditions for Eqs. (26). From the condition that system (26) be consistent at $t = 0$, we find that if one of the modes p, q, or k is absent in the set Ξ of initially excited modes, i.e., the amplitude of this mode equals zero at the initial time instant, its phase is $\pi/2$ and not arbitrary. Eventually, the initial conditions for set (26) can be written in compact form

$$a_j^{(1)}(0) = \delta_{i,j}h_j/2; \quad b_j^{(1)}(0) = \pm(1 - \delta_{i,j})\pi/2; \quad (27)$$
$$i \in \Xi; \quad j = k, p, q.$$

First-order coefficients in expansion (16) for resonantly coupled modes p, q, and k will be written as (see (18))

$$M_k^{(1)}(t) = 2a_k^{(1)}(\varepsilon t)\cos[(\omega_p - \omega_q)t - \beta_k^{(1)}(\varepsilon t)];$$
$$M_p^{(1)}(t) = 2a_p^{(1)}(\varepsilon t)\cos[\omega_p t + b_p^{(1)}(\varepsilon t)]; \quad (28)$$
$$M_q^{(1)}(t) = 2a_q^{(1)}(\varepsilon t)\cos[\omega_q t + b_q^{(1)}(\varepsilon t)],$$

where the coefficients $a_k^{(1)}$, $\beta_k^{(1)}$, $a_p^{(1)}$, $b_p^{(1)}$, $a_q^{(1)}$, and $b_q^{(1)}$ are the solutions to set (26) with initial conditions (27).

Note that in the second-order approximation, the resonant interaction between three modes is observed only if at least two of them are present in the initially excited spectrum, i.e., if their amplitudes are other than zero at $t = 0$. The third mode, even with a zero initial amplitude, will appear in the first-order vibration spectrum if its number meets the conditions that (for the case $p, q \in \Xi$, $k \notin \Xi$) $p + q + k$ is even and $|p - q| \le k \le p + q$, which follow from the requirement that the coefficients $\Lambda_{pqk}^{(-)}$, $\Lambda_{kqp}^{(+)}$, and $\Lambda_{pkq}^{(-)}$ in (26) be other than zero.

For $\varepsilon_d = 5$ and $\varepsilon = 0.3$, the time evolution of the first-order amplitudes of the third, sixth, and seventh modes (with $W = 3.361$) is shown in Fig. 1a and for the third, fourth, and fifth modes (with $W = 4.355$), in Fig. 1b. It is seen that the third mode, which was absent in the initial disturbance spectrum, is excited by resonantly taking up the energy from the highest mode. Calculations indicate that the variation of the permittivity ε_d affects resonant mode coupling insignificantly.

Prior to presenting solutions to Eq. (21), we note that for second-order calculations, it is sufficient to find the dependence of the coefficients $M_n^{(2)}(t)$ on the time T_0. Indeed, the neglect of the $M_n^{(2)}(t)$ vs. T_1 dependence (T_1 is the slow time) causes an error $O(\varepsilon t)$. In expansion (14) for $\xi(\Theta, t)$, this error produces an error $O(\varepsilon^3 t)$ and, hence, does not influence the accuracy of the second-order approximation within time intervals t $O(1)$. Because of this, when solving Eq. (21), we will assume that all the functions in its right-hand side do not depend on T_1 and equal their initial values.

Assuming that the partial solution to inhomogeneous equation (21) depends on T_0 in the same way as its right-hand side and adding this solution to the general solution to the corresponding homogeneous equation, we write the general solution to inhomogeneous equation (21) for the case $n \geq 2$; $n \neq p, q, k$ in the form

$$M_n^{(2)}(T_0) = a_n^{(2)} \exp[i(\omega_n T_0 + b_n^{(2)})] + \sum_{l=2}^{\infty} \sum_{m=2}^{\infty} \{\lambda_{lmn}^{(+)} A_l^{(1)}(0) A_m^{(1)}(0) \exp[i(\omega_l + \omega_m) T_0] + \lambda_{lmn}^{(-)} A_l^{(1)}(0) \overline{A_m^{(1)}(0)} \exp[i(\omega_l - \omega_m) T_0]\} + \text{c.c.} \quad (29)$$

Here, $a_n^{(2)}$ and $b_n^{(2)}$ are real constants that are determined from the initial conditions and the expressions for $\lambda_{lmn}^{(\pm)}$ are given in Appendix B.

In the second-order approximation, initial conditions (12) take the form

$$M_n^{(2)}(0) = -\xi_0 \delta_{n,0} - \xi_1 \delta_{n,1}; \quad \frac{\partial M_n^{(2)}(0)}{\partial T_0} + \frac{\partial M_n^{(1)}(0)}{\partial T_1} = 0. \quad (30)$$

Substituting (29) and (30) into (12), we arrive at the final expression for the second-order coefficients in expansion (16):

$$M_n^{(2)}(t) = \sum_{i \in \Xi} \sum_{j \in \Xi} h_i h_j \left\{ \lambda_{ijn}^{(+)} \sin\left[\frac{1}{2}(\omega_n + \omega_i + \omega_j)t\right] \times \sin\left[\frac{1}{2}(\omega_n - \omega_i - \omega_j)t\right] + \lambda_{ijn}^{(-)} \sin\left[\frac{1}{2}(\omega_n + \omega_i - \omega_j)t\right] \times \sin\left[\frac{1}{2}(\omega_n - \omega_i + \omega_j)t\right] \right\}; \quad n \geq 2; \quad n \neq p, q, k. \quad (31)$$

The time evolution of the zero ($n = 0$) and first ($n = 1$) modes is defined by requirements (10) and (11). With (10) and (11) written in the second-order approximation, we easily obtain

$$M_0^{(2)}(t) = -\frac{1}{2} \sum_{i \in \Xi} \frac{h_i^2}{(2i+1)} (1 + \cos(2\omega_i t)); \quad M_1^{(2)}(t) = -\sum_{i \in \Xi} \frac{9 i h_{i-1} h_i}{(2i-1)(2i+1)} \cos(\omega_i t) \cos(\omega_{i-1} t). \quad (32)$$

It should be noted that expressions (32) are consistent with general equation (21), which can be checked by direct substitution.

When solving (21) with $n = p$, q, or k, one should bear in mind that "dangerous" terms are excluded from the double sum on the right of (21) by virtue of relationships (24). We take into account this circumstance through the definition

$$D_{lm}^{pq} \equiv (1 - \delta_{l,p}\delta_{m,q})(1 - \delta_{l,q}\delta_{m,p}) \quad (33)$$

and write the general solutions for $n = p$, q, or k in a form similar to (29) but with additional coefficients: D_{lm}^{pq} before $\lambda_{lmk}^{(-)}$ (if $n = k$), D_{lm}^{kq} before $\lambda_{lmp}^{(+)}$ (if $n = p$), and D_{lm}^{kp} before $\lambda_{lmp}^{(-)}$ (if $n = q$).

These general solutions contain the real constants $a_k^{(2)}$, $\beta_k^{(2)}$, $a_p^{(2)}$, $b_p^{(2)}$, $a_q^{(2)}$, and $b_q^{(2)}$ that are specified by initial conditions (30).

If three resonantly coupled modes are present in the spectrum of initially excited modes, i.e., $(k, p, q) \in \Xi$, then, satisfying initial conditions (30), we obtain expressions for the time-dependent second-order coefficients in expansion (16) that are similar to (31):

$$M_k^{(2)}(t) = \sum_{i \in \Xi} \sum_{j \in \Xi} h_i h_j \left\{ \lambda_{ijk}^{(+)} \sin\left[\frac{1}{2}(\omega_p - \omega_q + \omega_i + \omega_j)t\right] \times \sin\left[\frac{1}{2}(\omega_p - \omega_q - \omega_i - \omega_j)t\right] \right.$$

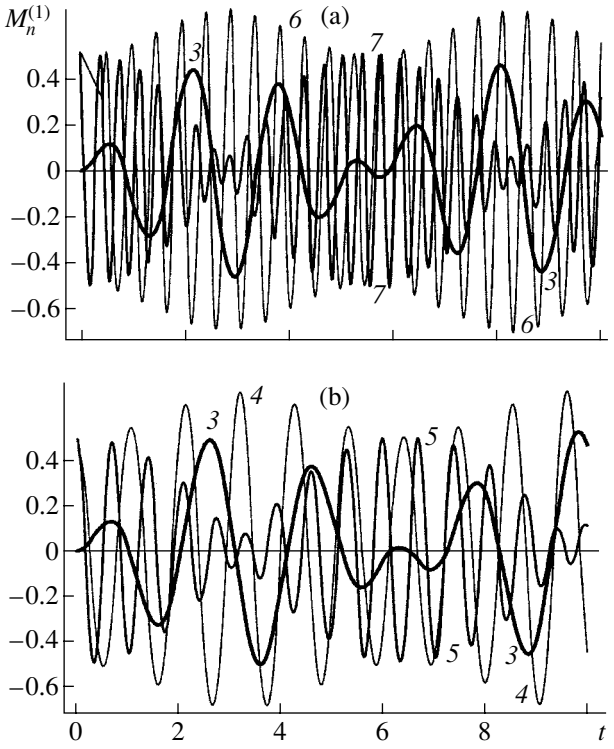


Fig. 1. Amplitudes $M_n^{(1)}$ of the resonantly interacting (a) third, sixth, and seventh modes and (b) third, fourth, and fifth modes vs. dimensionless time t. The figure by the curves indicates the mode number. Different modes are depicted by lines of different thickness.

$$+ D_{ij}^{pq}\lambda_{ijk}^{(-)}\sin\left[\frac{1}{2}(\omega_p - \omega_q + \omega_i - \omega_j)t\right]$$

$$\times \sin\left[\frac{1}{2}(\omega_p - \omega_q - \omega_i + \omega_j)t\right]\Bigg\};$$

$$M_p^{(2)}(t) = \sum_{i\in\Xi}\sum_{j\in\Xi} h_i h_j\Bigg\{D_{ij}^{kq}\lambda_{ijp}^{(+)}\sin\left[\frac{1}{2}(\omega_p + \omega_i + \omega_j)t\right]$$

$$\times \sin\left[\frac{1}{2}(\omega_p - \omega_i - \omega_j)t\right] \qquad (34)$$

$$+ \lambda_{ijp}^{(-)}\sin\left[\frac{1}{2}(\omega_p + \omega_i - \omega_j)t\right]\sin\left[\frac{1}{2}(\omega_p - \omega_i + \omega_j)t\right]\Bigg\};$$

$$M_q^{(2)}(t) = \sum_{i\in\Xi}\sum_{j\in\Xi} h_i h_j\Bigg\{\lambda_{ijq}^{(+)}\sin\left[\frac{1}{2}(\omega_q + \omega_i + \omega_j)t\right]$$

$$\times \sin\left[\frac{1}{2}(\omega_q - \omega_i - \omega_j)t\right]$$

$$+ D_{ij}^{kp}\lambda_{ijq}^{(-)}\sin\left[\frac{1}{2}(\omega_q + \omega_i - \omega_j)t\right]$$

$$\times \sin\left[\frac{1}{2}(\omega_q - \omega_i + \omega_j)t\right]\Bigg\}.$$

If one of the resonantly coupled modes has a zero amplitude and the amplitudes of the remaining two are other than zero, for example, $[p, q] \in \Xi$, $k \notin \Xi$, the coefficients $M_p^{(2)}(t)$ and $M_q^{(2)}(t)$ are given by (34), while for the kth mode, we have

$$M_k^{(2)}(t) = 2a_k^{(2)}\cos[(\omega_p - \omega_q)t - \beta_k^{(2)}]$$

$$+ \frac{1}{2}\sum_{i\in\Xi}\sum_{j\in\Xi} h_i h_j\{\lambda_{ijk}^{(+)}\cos[(\omega_i + \omega_j)t]$$

$$+ D_{ij}^{pq}\lambda_{ijl}^{(-)}\cos[(\omega_i - \omega_j)t]\};$$

$$a_k^{(2)} = ((\alpha_k^{(1)})^2 + (\alpha_k^{(2)}))^{21/2}; \quad \tan\beta_k^{(2)} = -\frac{\alpha_k^{(2)}}{\alpha_k^{(1)}}; \qquad (35)$$

$$\alpha_k^{(1)} \equiv \frac{1}{4}\sum_{i\in\Xi}\sum_{j\in\Xi} h_i h_j(\lambda_{ijk}^{(+)} + D_{ij}^{pq}\lambda_{ijk}^{(-)});$$

$$\alpha_k^{(2)} \equiv \frac{h_p h_q}{8(\omega_p - \omega_q)^2}\Lambda_{pqk}^{(-)}.$$

Similar expressions can also be written for two other possible situations: $[k, q] \in \Xi$, $p \notin \Xi$ and $[k, p] \in \Xi$, $q \notin \Xi$. In the former case,

$$M_p^{(2)}(t) = 2a_p^{(2)}\cos[\omega_p t + b_p^{(2)}]$$

$$+ \frac{1}{2}\sum_{i\in\Xi}\sum_{j\in\Xi} h_i h_j\{D_{ij}^{kq}\lambda_{ijp}^{(+)}\cos[(\omega_i + \omega_j)t] \qquad (36)$$

$$+ \lambda_{ijp}^{(-)}\cos[(\omega_i - \omega_j)t]\}.$$

In the latter,

$$M_q^{(2)}(t) = 2a_q^{(2)}\cos[\omega_p t + b_q^{(2)}]$$

$$+ \frac{1}{2}\sum_{i\in\Xi}\sum_{j\in\Xi} h_i h_j\{\lambda_{ijp}^{(+)}\cos[(\omega_i + \omega_j)t]$$

$$+ D_{ij}^{kp}\lambda_{ijp}^{(-)}\cos[(\omega_i - \omega_j)t]\};$$

$$a_m^{(2)} = ((\alpha_m^{(1)})^2 + (\alpha_m^{(2)})^2)^{1/2};$$

$$\tan b_m^{(2)} = -\frac{\alpha_m^{(2)}}{\alpha_m^{(1)}}; \quad m = (p, q); \qquad (37)$$

$$\alpha_p^{(1)} \equiv \frac{1}{4}\sum_{i\in\Xi}^{\infty}\sum_{j\in\Xi}^{\infty} h_i h_j(D_{ij}^{kq}\lambda_{ijp}^{(+)} + \lambda_{ijp}^{(-)});$$

$$\alpha_p^{(2)} \equiv \frac{h_k h_q}{8\omega_p^2}\Lambda_{kqp}^{(+)};$$

$$\alpha_q^{(1)} \equiv \frac{1}{4}\sum_{i\in\Xi}^{\infty}\sum_{j\in\Xi}^{\infty} h_i h_j(\lambda_{ijq}^{(+)} + D_{ij}^{kp}\lambda_{ijq}^{(-)});$$

$$\alpha_q^{(2)} \equiv \frac{h_k h_p}{8\omega_q^2}\Lambda_{pkq}^{(-)}.$$

The coefficients omitted here are given by (34) in both cases.

Figure 2 shows the time dependences of the third mode amplitude for $\varepsilon_d = 5$, $\varepsilon = 0.3$, and $W = 4.355$ accurate to the first and second order of smallness for the resonant interaction between the third, fourth, and fifth modes. It is easy to check that the second-order correction to the amplitude is small. Yet it should be recalled once more that resonant coupling between modes of capillary drop vibration is a nonlinear effect that can be "observed" only in the second-order approximation in the initial disturbance amplitude.

(6) Consider separately the case when only two modes interact resonantly; that is, when the relationship

$$\omega_k = 2\omega_p$$

takes place.

The same analysis as before yields for the first-order time-dependent coefficients in expansion (16)

$$M_k^{(1)}(t) = 2a_k^{(1)}(\varepsilon t)\cos[2\omega_k t - \beta_k^{(1)}(\varepsilon t)];$$
$$M_p^{(1)}(t) = 2a_p^{(1)}(\varepsilon t)\cos[\omega_p t + b_p^{(1)}(\varepsilon t)], \quad (38)$$

where the real functions $a_k^{(1)}(\varepsilon t)$, $\beta_k^{(1)}(\varepsilon t)$, $a_p^{(1)}(\varepsilon t)$, and $b_p^{(1)}(\varepsilon t)$ are the solutions to the set of differential equations

$$4\omega_k \frac{da_k^{(1)}(T_1)}{dT_1} = \Lambda_{ppk}^{(+)}[a_p^{(1)}(T_1)]^2 \sin[\varphi_{kp}^{(1)}(T_1)];$$

$$4\omega_k a_k^{(1)}(T_1)\frac{d\beta_k^{(1)}(T_1)}{dT_1} = 4\omega_k a_k^{(1)}(T_1)\sigma + \Lambda_{ppk}^{(+)}[a_p^{(1)}(T_1)]^2 \cos[\varphi_{kp}^{(1)}(T_1)];$$

$$2\omega_p \frac{da_p^{(1)}(T_1)}{dT_1} = -\Lambda_{kpp}^{(-)} a_k^{(1)}(T_1) a_p^{(1)}(T_1)\sin[\varphi_{kp}^{(1)}(T_1)];$$

$$2\omega_p a_p^{(1)}(T_1)\frac{db_p^{(1)}(T_1)}{dT_1} = -\Lambda_{kpp}^{(1)} a_k^{(1)}(T_1) a_p^{(1)}(T_1)\cos[\varphi_{kp}^{(1)}(T_1)]; \quad (39)$$

$$\varphi_{kp}^{(1)}(T_1) = \beta_k^{(1)}(T_1) + 2b_p^{(1)}(T_1).$$

Relationships (20) imply the following sets of the initial conditions for system (39):

$$[k, p] \in \Xi: a_k^{(1)}(0) = h_k/2; \quad \beta_k^{(1)}(0) = 0;$$
$$a_p^{(1)}(0) = h_p/2, \quad b_p^{(1)}(0) = 0;$$
$$k \notin \Xi, \quad p \in \Xi: a_k^{(1)}(0) = 0; \quad \beta_k^{(1)}(0) = \pi/2;$$
$$a_p^{(1)}(0) = h_p/2; \quad b_p^{(1)}(0) = 0.$$

In the case $p \notin \Xi$, $k \in \Xi$ (i.e., $a_p^{(1)}(0) = 0$, $a_k^{(1)}(0) = h_k/2$), resonant coupling between the modes p and k is absent in this approximation, since from (39) at $t = 0$ we have

$$\frac{da_k^{(1)}(0)}{dT_1} = \frac{da_p^{(1)}(0)}{dT_1} = 0,$$

i.e., the amplitudes $a_p^{(1)}$ and $a_k^{(1)}$ retain their initial values.

Figure 3 depicts the time dependences of the amplitudes of the resonantly interacting fourth and sixth modes with $\varepsilon_d = 5$, $\varepsilon = 0.3$, and $W = 3.248$ for two initial conditions: (1) both modes are present in the initial disturbance spectrum and (2) initially the fourth mode is excited and the six mode has a zero amplitude. If initially only the sixth mode is present, the resonant build-up of the fourth mode does not occur [8]. As was mentioned above, this resonance was found in [1]; however, it is not the only two-mode resonance at subcritical values of the parameter W. For example, for $\varepsilon_d = 5$ and $W = 3.607$, the resonance between the ninth and fourteenth modes takes place ($\omega_{14} = 2\omega_9$).

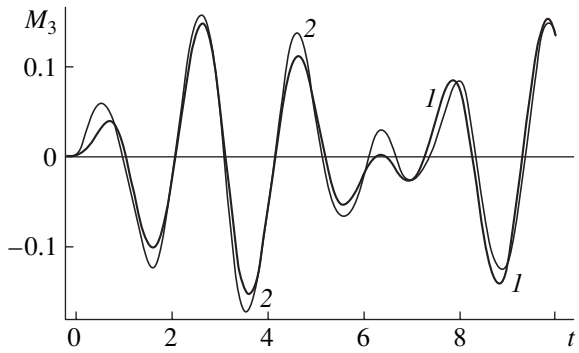


Fig. 2. Variation of the third mode with time calculated in the (*1*) first and (*2*) second approximations.

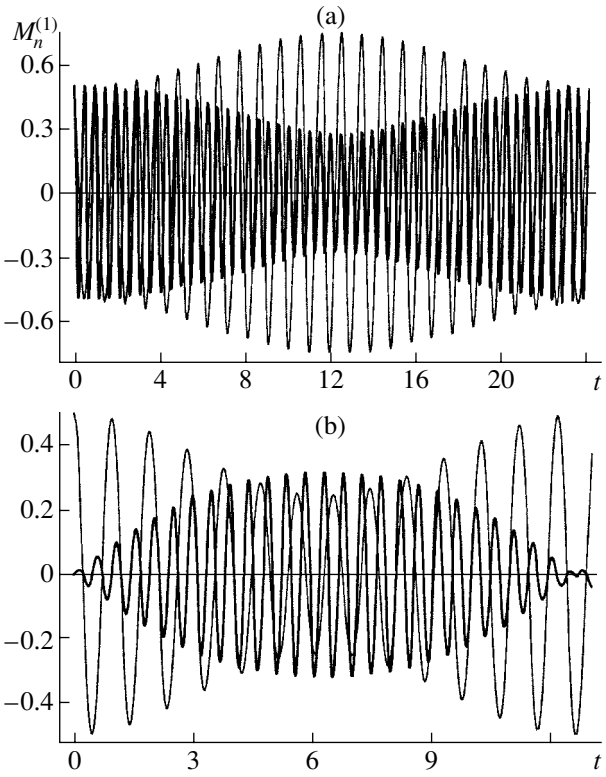


Fig. 3. The same as in Fig. 1 for two-mode coupling: (a) the fourth and sixth modes are initially excited and (b) the fourth mode alone is initially excited. Fine line, fourth mode; thick line, sixth mode.

In the second-order approximation for two-mode resonance when $[k, p] \in \Xi$, we have

$$M_k^{(2)}(t) = \sum_{i \in \Xi} \sum_{j \in \Xi} h_i h_j \left\{ D_{ij}^{pp} \lambda_{ijk}^{(+)} \sin\left[\frac{1}{2}(2\omega_p + \omega_i + \omega_j)t\right] \right.$$

$$\times \sin\left[\frac{1}{2}(2\omega_p - \omega_i - \omega_j)t\right] + \lambda_{ijk}^{(-)} \sin\left[\frac{1}{2}(2\omega_p + \omega_i - \omega_j)t\right]$$

$$\times \sin\left[\frac{1}{2}(2\omega_p - \omega_i + \omega_j)t\right]\Bigg\}; \tag{40}$$

$$M_p^{(2)}(t) = \sum_{i \in \Xi} \sum_{j \in \Xi} h_i h_j \Bigg\{ \lambda_{ijp}^{(+)} \sin\left[\frac{1}{2}(\omega_p + \omega_i + \omega_j)t\right]$$

$$\times \sin\left[\frac{1}{2}(\omega_p - \omega_i - \omega_j)t\right] + D_{ij}^{kp} \lambda_{ijp}^{(-)} \sin\left[\frac{1}{2}(\omega_p + \omega_i - \omega_j)t\right]$$

$$\times \sin\left[\frac{1}{2}(\omega_p - \omega_i + \omega_j)t\right]\Bigg\}.$$

If the mode k is absent in the initial disturbance spectrum ($k \notin \Xi$, $p \in \Xi$), the second-order time-dependent coefficient in expansion (16) for this mode has the form (instead of (40))

$$M_k^{(2)}(t) = 2a_k^{(2)} \cos[2\omega_p(t) - \beta_k^{(2)}]$$

$$+ \frac{1}{2} \sum_{i \in \Xi} \sum_{j \in \Xi} h_i h_j \{ D_{ij}^{pp} \lambda_{ijk}^{(+)} \sin[(\omega_i + \omega_j)t]$$

$$+ \lambda_{ijk}^{(-)} \cos[(\omega_i - \omega_j)t] \};$$

$$a_k^{(2)} = ((\alpha_k^{(1)})^2 + (\alpha_k^{(2)})^2)^{1/2}; \quad \tan\beta_k^{(2)} = -\frac{\alpha_k^{(2)}}{\alpha_k^{(1)}}; \tag{41}$$

$$\alpha_k^{(1)} \equiv \frac{1}{4} \sum_{i \in \Xi} \sum_{j \in \Xi} h_i h_j (D_{ij}^{pp} \lambda_{ijk}^{(+)} + \lambda_{ijk}^{(-)});$$

$$\alpha_k^{(2)} \equiv \frac{h_p^2}{64\omega_p^2} \Lambda_{ppk}^{(+)}.$$

(7) The final expression for the shape of the drop surface accurate to second-order terms in the initial disturbance amplitude ε is given by (see (1), (14), and (16))

$$r(\Theta, t) \approx 1 + \varepsilon \Bigg\{ \sum_{i \in \mathcal{R}} M_i^{(1)} P_i(\mu) \Bigg\}$$

$$+ \varepsilon^2 \Bigg\{ \sum_{n=0}^{\infty} M_n^{(2)}(t) P_n(\mu) \Bigg\} + O(\varepsilon^3), \tag{42}$$

where $\mathcal{R}$ is the set of the numbers of capillary vibration modes whose amplitudes are of a second order of smallness in ε. This set combines the set Ξ of the numbers of modes excited initially (at $t = 0$) (see (12)) and the numbers of modes generated by resonant mode interaction. In (42), the amplitudes $M_i^{(1)}(t)$ are given by formulas (19) and (23) or, in the presence of the resonance, by (28) or (38). The amplitudes $M_n^{(2)}(t)$ are given by expressions (31) and (32) or, in the presence of resonance, by (34)–(37) or (40) and (41) according to initial conditions and type of resonance coupling. Thus, in the approximation linear in ε, the shape of the drop is defined by the vibrations of modes present in the initial perturbation spectrum and of modes excited by resonance interaction. The vibration eigenfrequencies ω_n depend on the permittivity ε_d of the liquid (see (17)) and grow as ε_d diminishes. The critical value of the parameter W (which is related to the charge Q of the drop) at which the surface of the drop loses stability is determined from the vanishing condition for the fundamental mode frequency ($n = 2$), $W_{cr} = 0.25\kappa_1(2, \varepsilon_d)$, and also grows with decreasing ε_d. Thus, a volumetrically charged dielectric drop is more stable against its charge than a drop of perfectly conducting surface-charged liquid.

Note that the second-order amplitudes $M_n^{(2)}(t)$ are proportional to the coefficients $\lambda_{ijn}^{(\pm)}$, which, in turn, are expressed through the Clebsch–Gordan coefficients [13] (see Appendix B). This means that $\lambda_{ijn}^{(\pm)}$ are nonzero only if the indices satisfy the relationships

$$|i - j| \le n \le i + j; \quad i + j + n = 2g, \tag{43}$$

where g is an integer.

Therefore, the number of modes excited by second-order mode coupling is specified by conditions (43), i.e., by the fact that the coefficients $\lambda_{iin}^{(\pm)}$ and $\lambda_{jjn}^{(\pm)}$ in expressions (31) are other than zero only for even n from the intervals $[0, 2i]$ and $[0, 2j]$, respectively, and $\lambda_{ijn}^{(\pm)}$ and $\lambda_{jin}^{(\pm)}$, for n from the interval $[|i - j|, (i + j)]$ such that $n + i + j$ is even. Therefore, when two, even and odd, modes (with numbers p and q) are excited simultaneously, their coupling will generate all even modes from the interval $[0, \max(2p, 2q)]$ and all odd modes from the interval $[|p - q|, (p + q)]$. If two odd or even modes are initially excited, their resonance will generate only even modes from the interval $[0, \max(2p, 2q)]$ in the second-order approximation.

As was mentioned above, this is also true for secondary resonances. If one of the resonantly coupled modes obeying (22) (say, the mode k) is absent in the initially excited spectrum ($k \notin \Xi$), its resonant excitation is possible if the numbers of the interacting modes k, p, and q meet conditions like (43):

$$|p - q| \le k \le p + q; \quad p + q + k = 2g.$$

From (32), it also follows that if the spectrum describing the initial disturbance of the surface contains two (even and odd) neighboring vibration modes, the amplitude of the former will be different from zero in the second-order approximation. This provides the center-of-mass immobility at $t \geq 0$ and its position at the origin by compensating for that displacement of the center of mass caused by the excitation of two neighboring vibration modes.

CONCLUSION

The study of the nonlinear surface dynamics of a nonviscous volumetrically charged dielectric spherical drop subjected to an arbitrary initial deformation shows that the number of resonances due to three-mode coupling is infinitely large if the charge of the drop is subcritical. Resonant mode coupling causes energy transfer from higher to lower modes. The amplitudes of lower modes increase, and the drop may become unstable against the self-charge. The third mode is the lowest among those excited as a result of resonant coupling with higher modes. Although nonlinear mode coupling is a second-order effect, it leads to an increase in the amplitude of an excited mode up to a first-order quantity (even if this mode had a zero amplitude).

A possible mechanism behind the disintegration of a charged drop in the case when the self-charge is small (subcritical in terms of linear instability) may be associated with the resonance transfer of the capillary vibration energy from higher to lower modes.

APPENDIX A: EXPRESSION FOR THE PRESSURE OF THE ELECTRIC FIELD OF THE DROP SELF-CHARGE

The problem of finding the internal and external electrostatic potentials $\Phi_{in}(\mathbf{r}, t)$ and $\Phi_{ex}(\mathbf{r}, t)$, which are needed to express the pressure on the drop surface in dynamic boundary condition (7), involves Eqs. (3) and boundary conditions (5), (8), and (9). Substituting expansion (14) into them and requiring these equations to be fulfilled for each order of smallness m, we arrive at a boundary-value problem for the functions $\Phi_{in}^{(m)}$ and $\Phi_{ex}^{(m)}$, where $m = 0, 1, 2$.

Since, a solution obviously possesses central symmetry in the zeroth approximation, the Laplace and Poisson equation for $\Phi_{in}^{(0)}$ and $\Phi_{ex}^{(0)}$, respectively, in (3) turn into ordinary second-order differential equations whose solutions are found in the form $\sim Cr^s$. With boundary conditions (8) and (9) satisfied in the zeroth-order approximation;

$$r = 1: \Phi_{in}^{(0)} = \Phi_{ex}^{(0)}; \quad \varepsilon_d \frac{d\Phi_{in}^{(0)}}{dr} = \frac{d\Phi_{ex}^{(0)}}{dr},$$

we find

$$\Phi_{ex}^{(0)} = \frac{Q}{r} + \Phi_\infty; \quad \Phi_{in}^{(0)} = \frac{Q}{2\varepsilon}(2\varepsilon_d + 1 - r^2) + \Phi_\infty, \quad \text{(A1)}$$

where Φ_∞ is a constant that equals the adopted value of the external potential at infinity.

The first- and second-order functions $\Phi_{in}^{(m)}$ and $\Phi_{ex}^{(m)}$ ($m = 1, 2$) are the solutions to the Laplace equation and, in view of boundary conditions (5), can be written in the form

$$\Phi_{ex}^{(m)}(r, \Theta, T_0, T_1) = \sum_{n=0}^{\infty} F_n^{(m, ex)}(T_0, T_1) r^{-(n+1)} P_n(\mu);$$

$$\Phi_{in}^{(m)}(r, \Theta, T_0, T_1) = \sum_{n=0}^{\infty} F_n^{(m, in)}(T_0, T_1) r^n P_n(\mu); \quad \text{(A2)}$$

$$m = 1; 2.$$

The functions $\xi^{(m)}(\Theta, T_0, T_1)$ ($m = 1, 2$) in expansion (14), which describe the shape of the drop in the first and second approximations are also represented as expansions in Legendre polynomials:

$$\xi^{(m)}(\Theta, T_0, T_1) = \sum_{n=0}^{\infty} M_n^{(m)}(T_0, T_1) P_n(\mu). \quad \text{(A3)}$$

The unknown functions of time $F_n^{(m, ex)}(T_0, T_1)$ and $F_n^{(m, in)}(T_0, T_1)$ in (A2) can be expressed through the amplitudes $M_n^{(m)}(T_0, T_1)$ in (A3) with conditions (8) and (9) on the free surface. In the first-order approximation, these conditions take the form

$$r = 1: \Phi_{in}^{(1)} + \frac{d\Phi_{in}^{(0)}}{dr}\xi^{(1)} = \Phi_{ex}^{(1)} + \frac{d\Phi_{ex}^{(0)}}{dr}\xi^{(1)};$$

$$\varepsilon_d\left[\frac{\partial\Phi_{in}^{(1)}}{\partial r} + \frac{d^2\Phi_{in}^{(0)}}{dr^2}\xi^{(1)}\right] = \frac{\partial\Phi_{ex}^{(1)}}{\partial r} + \frac{d^2\Phi_{ex}^{(0)}}{dr^2}\xi^{(1)}.$$

Substituting solutions (A2) into them with $m = 1$ yields

$$\begin{aligned} F_n^{(1, ex)}(T_0, T_1) &= \kappa_1^{(ex)}(n, \varepsilon_d) Q M_n^{(1)}(T_0, T_1); \\ F_n^{(1, in)}(T_0, T_1) &= -\kappa_1^{(in)}(n, \varepsilon_d) Q M_n^{(1)}(T_0, T_1). \end{aligned} \quad \text{(A4)}$$

The expressions for the coefficients $\kappa_1^{(ex)}(n, \varepsilon_d)$ and $\kappa_1^{(in)}(n, \varepsilon_d)$ are given in Appendix B.

In the second approximation, conditions (8) and (9) have the form

$$r = 1: \Phi_{in}^{(2)} + \frac{d\Phi_{in}^{(0)}}{dr}\xi^{(2)} + \frac{\partial\Phi_{in}^{(1)}}{\partial r}\xi^{(1)} + \frac{1}{2}\frac{d^2\Phi_{in}^{(0)}}{dr^2}(\xi^{(1)})^2$$
$$= \Phi_{ex}^{(2)} + \frac{d\Phi_{ex}^{(0)}}{dr}\xi^{(2)} + \frac{\partial\Phi_{ex}^{(1)}}{\partial r}\xi^{(1)} + \frac{1}{2}\frac{d^2\Phi_{ex}^{(0)}}{dr^2}(\xi^{(1)})^2;$$

$$\varepsilon_d\left[\frac{d\Phi_{in}^{(2)}}{dr} + \frac{d^2\Phi_{in}^{(0)}}{dr^2}\xi^{(2)} + \frac{d^2\Phi_{in}^{(1)}}{dr^2}\xi^{(1)} + \frac{1}{2}\frac{d^3\Phi_{in}^{(0)}}{dr^3}(\xi^{(1)})^2\right.$$
$$\left. - \frac{1}{2}\frac{d\Phi_{in}^{(0)}}{dr}\left(\frac{\partial\xi^{(1)}}{\partial\Theta}\right)^2 - \frac{\partial\Phi_{in}^{(1)}}{\partial\Theta}\frac{\partial\xi^{(1)}}{\partial\Theta}\right] = \frac{\partial\Phi_{ex}^{(2)}}{\partial r}$$
$$+ \frac{d^2\Phi_{ex}^{(0)}}{dr^2}\xi^{(2)} + \frac{\partial^2\Phi_{ex}^{(1)}}{\partial r^2}\xi^{(1)} + \frac{1}{2}\frac{d^3\Phi_{ex}^{(0)}}{dr^3}(\xi^{(1)})^2$$
$$- \frac{1}{2}\frac{d\Phi_{ex}^{(0)}}{dr}\left(\frac{\partial\xi^{(1)}}{\partial\Theta}\right)^2 - \frac{\partial\Phi_{ex}^{(1)}}{\partial\Theta}\frac{\partial\xi^{(1)}}{\partial\Theta}.$$

Then, for the functions $F_n^{(2, ex)}$ and $F_n^{(2, in)}$ in solutions (A2), one obtains

$$F_n^{(2, ex)}(T_0, T_1) = \frac{(n\varepsilon_d - n + 3)}{(n\varepsilon_d + n + 1)}QM_n^{(2)}(T_0, T_1)$$
$$+ \sum_{i=2}^{\infty}\sum_{j=2}^{\infty}[K_{ijn}\kappa_2^{(ex)}(i, n, \varepsilon_d) - \alpha_{ijn}\kappa_2(i, n, \varepsilon_d)]$$
$$\times QM_i^{(1)}(T_0, T_1)M_j^{(1)}(T_0, T_1);$$
$$F_n^{(2, in)}(T_0, T_1) = -\frac{[(n-2)\varepsilon_d - n - 1]}{\varepsilon_d(n\varepsilon_d + n + 1)}QM_n^{(2)}(T_0, T_1) \quad \text{(A5)}$$
$$+ \sum_{i=2}^{\infty}\sum_{j=2}^{\infty}[K_{ijn}\kappa_2^{(in)}(i, n, \varepsilon_d) - \alpha_{ijn}\kappa_2(i, n, \varepsilon_d)]$$
$$\times QM_i^{(1)}(T_0, T_1)M_j^{(1)}(T_0, T_1).$$

All necessary designations are given in Appendix B.

The electric field pressure p_E on the drop is the sum of the following terms in dynamic boundary condition (7):

$$p_E = \left[-\mu_q\Phi_{in} + (\varepsilon_d - 1)\frac{1}{8\pi}(\nabla\Phi_{ex})^2\right.$$
$$\left. - \frac{(\varepsilon_d - 1)^2}{\varepsilon_d}\frac{1}{8\pi}(\mathbf{n}\cdot\nabla\Phi_{ex})^2\right]\Big|_{r=1+\xi}. \quad \text{(A6)}$$

In (A6), the unit normal vector $\mathbf{n}$ to the surface whose shape is specified by Eq. (1) is given by

$$\mathbf{n} \approx \mathbf{e}_r\left[1 - \frac{\varepsilon^2}{2}\left(\frac{\partial\xi^{(1)}}{\partial\Theta}\right)^2\right]$$
$$- \mathbf{e}_\Theta\left[\varepsilon\frac{\partial\xi^{(1)}}{\partial\Theta} + \varepsilon^2\left(\frac{\partial\xi^{(2)}}{\partial\Theta} - \varepsilon^{(1)}\frac{\partial\xi^{(1)}}{\partial\Theta}\right)\right] \quad \text{(A7)}$$

up to second-order smalls.

Expanding (A6) in the small parameter ε in view of representations (14) yields for (A7)

$$p_E \approx \left\{-\frac{3Q}{4\pi}\Phi_{in}^{(0)} + \frac{(\varepsilon_d - 1)}{\varepsilon_d}\frac{1}{8\pi}\left(\frac{d\Phi_{ex}^{(0)}}{dr}\right)^2\right\}_{r=1}$$
$$+ \varepsilon\left\{-\frac{3Q}{4\pi}\left[\Phi_{in}^{(1)} + \frac{d\Phi_{in}^{(0)}}{dr}\xi^{(1)}\right] + \frac{(\varepsilon_d - 1)}{\varepsilon_d}\frac{1}{4\pi}\right.$$
$$\left.\times\frac{d\Phi_{ex}^{(0)}}{dr}\left[\frac{d\Phi_{ex}^{(1)}}{\partial r} + \frac{\partial^2\Phi_{ex}^{(0)}}{\partial r^2}\xi^{(1)}\right]\right\}_{r=1} + \varepsilon^2\left\{-\frac{3Q}{4\pi}\right.$$
$$\times\left[\Phi_{in}^{(2)} + \frac{d\Phi_{in}^{(0)}}{dr}\xi^{(2)} + \frac{\partial\Phi_{in}^{(1)}}{\partial r}\xi^{(1)} + \frac{1}{2}\frac{d^2\Phi_{in}^{(0)}}{dr^2}(\xi^{(1)})^2\right]$$
$$+ \frac{(\varepsilon_d - 1)}{\varepsilon_d}\frac{1}{8\pi}\left[2\frac{d\Phi_{ex}^{(0)}}{dr}\left(\frac{\partial\Phi_{ex}^{(2)}}{\partial r} + \frac{d^2\Phi_{ex}^{(0)}}{dr^2}\xi^{(2)}\right)\right.$$
$$+ 2\left(\frac{d^2\Phi_{ex}^{(0)}}{dr^2}\frac{\partial\Phi_{ex}^{(1)}}{\partial r} + \frac{d\Phi_{ex}^{(0)}}{dr}\frac{d^2\Phi_{ex}^{(1)}}{dr^2}\right)\xi^{(1)}$$
$$+ \left(\frac{d\Phi_{ex}^{(0)}}{dr^2}\frac{d^3\Phi_{ex}^{(0)}}{\partial r^3} + \frac{d^2\Phi_{ex}^{(0)}}{dr^2}\frac{\partial^2\Phi_{ex}^{(0)}}{dr^2}\right)(\xi^{(1)})^2$$
$$+ \left(\frac{\partial\Phi_{ex}^{(1)}}{\partial r}\right)^2 + \varepsilon_d\left(\frac{\partial\Phi_{ex}^{(1)}}{\partial\Theta}\right)^2 + (\varepsilon_d - 1)$$
$$\left.\left.\times\frac{\partial\Phi_{ex}^{(0)}}{\partial r}\left(2\frac{\partial\Phi_{ex}^{(1)}}{\partial\Theta} + \frac{\partial\Phi_{ex}^{(0)}}{dr}\frac{d\xi^{(1)}}{d\Theta}\right)\frac{\partial\xi^{(1)}}{\partial\Theta}\right\}\right|_{r=1} + O(\varepsilon^3).$$

Substituting solutions (A1)–(A5), we eventually find for the electric field pressure

$$p_E = -\mu_q(Q + \Phi_\infty) + \frac{Q^2}{8\pi}\frac{(\varepsilon_d - 1)}{\varepsilon_d} + \varepsilon_d\frac{Q^2}{8\pi}$$
$$\times\sum_{n=0}^{\infty}2(n-1)\kappa_1(n, \varepsilon_d)M_n^{(1)}(T_0, T_1)P_n(\mu)$$

$$+\varepsilon^2\frac{Q^2}{8\pi}\sum_{n=0}^{\infty}\Bigg\{2(n-1)\kappa_1(n,\varepsilon_d)M_n^{(2)}(T_0,T_1)$$

$$+\sum_{i=2}^{\infty}\sum_{j=2}^{\infty}[K_{ijn}\kappa_3(i,j,n,\varepsilon_d)+\alpha_{ijn}\kappa_4(i,j,n,\varepsilon_d)]$$

$$\times M_i^{(1)}(T_0,T_1)M_j^{(1)}(T_0,T_1)\Bigg\}P_n(\mu)+O(\varepsilon^3).$$

All definitions omitted here are given in Appendix B.

APPENDIX B: COEFFICIENTS APPEARING IN THE TIME DEPENDENCE OF THE SHAPE OF THE NONLINEARLY VIBRATING DROP (16)

$$\lambda_{ijn}^{(\pm)}\equiv[\gamma_{ijn}\pm\omega_i\omega_j\eta_{ijn}][\omega_n^2-(\omega_i\pm\omega_j)^2]^{-1};$$

$$\omega_n^2\equiv n(n-1)[(n+2)-W\kappa_1(n,\varepsilon_d)];\quad W\equiv\frac{Q^2}{4\pi};$$

$$\gamma_{ijn}\equiv K_{ijn}\Bigg[\omega_i^2(n-i+1)+2n[j(j+1)-1]$$

$$+\kappa_3(i,j,n,\varepsilon_d)n\frac{W}{2}\Bigg]+\alpha_{ijn}\Bigg[\frac{1}{i}\omega_i^2+\kappa_4(i,j,n,\varepsilon_d)n\frac{W}{2}\Bigg];$$

$$\eta_{ijn}\equiv K_{ijn}\left(\frac{n}{2}-i+1\right)+\alpha_{ijn}\frac{1}{i}\left(1+\frac{n}{2j}\right);$$

$$\kappa_1(n,\varepsilon_d)\equiv\frac{n\varepsilon_d^2-(2n-5)\varepsilon_d+(n+1)}{\varepsilon_d(n\varepsilon_d+n+1)};$$

$$\kappa_3(i,j,n,\varepsilon_d)\equiv\frac{1}{\varepsilon_d}\{2(n+1)(\varepsilon_d-1)\kappa_2^{(\mathrm{ex})}(i,n,\varepsilon_d)$$

$$+(i+1)(\varepsilon_d-1)\kappa_1^{(\mathrm{ex})}(i,\varepsilon_d)[(j+1)\kappa_1^{(\mathrm{ex})}(j,\varepsilon_d)$$

$$-2(i+4)]-6\varepsilon_d[\kappa_2^{(\mathrm{in})}(i,n,\varepsilon_d)-i\kappa_1^{(\mathrm{in})}(i,\varepsilon_d)]+10\varepsilon_d-7\};$$

$$\kappa_4(i,j,n,\varepsilon_d)\equiv\frac{1}{\varepsilon_d}\{2[n+1-(n-2)\varepsilon_d]\kappa_2(i,n,\varepsilon_d)$$

$$+(\varepsilon_d-1)[\kappa_1^{(\mathrm{ex})}(i,\varepsilon_d)(\varepsilon_d\kappa_1^{(\mathrm{ex})}(j,\varepsilon_d)-2(\varepsilon_d-1))+\varepsilon_d-1]\};$$

$$\kappa_1^{(\mathrm{in})}(n,\varepsilon_d)\equiv\frac{((n-2)\varepsilon_d-n-1)}{\varepsilon_d(n\varepsilon_d+n+1)};$$

$$\kappa_1^{(\mathrm{ex})}(n,\varepsilon_d)\equiv\frac{(n\varepsilon_d-n+3)}{(n\varepsilon_d+n+1)};$$

$$\kappa_2^{(\mathrm{in})}(i,n,\varepsilon_d)\equiv\frac{1}{(n\varepsilon_d+n+1)}\Bigg\{-(i+1)(n-i-1)$$

$$\times\kappa_1^{(\mathrm{ex})}(j,\varepsilon_d)+i((i+1)\varepsilon_d+n+1)$$

$$\times\kappa_1^{(\mathrm{ex})}(i,\varepsilon_d)+\frac{1}{2\varepsilon_d}[2(n-2)\varepsilon_d+n+1]\Bigg\};$$

$$\kappa_2^{(\mathrm{ex})}(i,n,\varepsilon_d)\equiv\frac{1}{(n\varepsilon_d+n+1)}\Bigg\{(i+1)(n\varepsilon_d+i+2$$

$$\times\kappa_1^{(\mathrm{ex})}(j,\varepsilon_d)+i(n-i+1)\varepsilon_d\kappa_1^{(\mathrm{in})}(i,\varepsilon_d)$$

$$-\frac{1}{2}2(n-2)\varepsilon_d+n+6)\Bigg\};$$

$$\kappa_2(i,n,\varepsilon_d)\equiv\frac{1}{(n\varepsilon_d+n+1)}\{\kappa_1^{(\mathrm{ex})}(i,\varepsilon_d)+\varepsilon_d\kappa_1^{(\mathrm{in})}(i,\varepsilon_d)\};$$

$$K_{ijn}\equiv[C_{i0j0}^{n0}]^2;$$

$$\alpha_{ijn}\equiv-\sqrt{i(i+1)j(j+1)}C_{i0j0}^{n0}C_{i(-1)j1}^{n0};$$

$$C_{i0j0}^{n0}\equiv\begin{cases}0, & \text{if}\quad i+j+n=2g+1,\\ \text{where } g \text{ is an integer;}\\ \dfrac{(-1)^{g-n}\sqrt{2n+1}g!}{(g-i)!(g-j)!(g-n)!}\\ \times\left[\dfrac{(2g-2i)!(2g-2j)!(2g-2n)!}{(2g+1)!}\right]^{1/2}, & \\ \text{if}\quad i+j+n=2g;\end{cases}$$

$$C_{i(-1)j1}^{n0}\equiv\sqrt{2n+1}n!$$

$$\times\left[\frac{(i+j-n)!i(i+1)}{(n+i-j)!(n-i+j)!(i+j+n+1)!j(j+1)}\right]^{1/2}$$

$$\times\sum_z\frac{(-1)^{i+1+z}(i+z-1)!(n+j-z+1)!}{z!(i-z+1)!(n-z)!(j-n+z-1)!}.$$

In the last sum, summation is over all integer z for which the expressions under the factorial sign are nonnegative, and C_{i0j0}^{n0} and $C_{i(-1)j1}^{n0}$ are the Clebsch–Gordan coefficients [13] that are other than zero only if the subscripts satisfy the relationships

$$|i-j|\le n\le(i+j);\quad i+j+n=2g,\ g \text{ is an integer.}$$

ACKNOWLEDGMENTS

This work was supported by grant no. 00-15-9925 of the President of the Russian Federation.

REFERENCES

1. J. A. Tsamopoulos and R. A. Brown, J. Fluid Mech. **147**, 373 (1984).
2. J. A. Tsamopoulos, T. R. Akylas, and R. A. Brown, Proc. R. Soc. London, Ser. A **401**, 67 (1985).

3. O. A. Basaran and L. E. Scriven, Phys. Fluids A **1**, 795 (1989).
4. R. Natarayan and R. A. Brown, Proc. R. Soc. London, Ser. A **410**, 209 (1987).
5. N. A. Pelekasis, J. A. Tsamopoulos, and G. D. Manolis, Phys. Fluids A **2**, 1328 (1990).
6. Z. C. Feng, J. Fluid Mech. **333**, 1 (1997).
7. D. F. Belonozhko and A. I. Grigor'ev, Zh. Tekh. Fiz. **70** (8), 45 (2000) [Tech. Phys. **45**, 1001 (2000)].
8. S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **26** (22), 76 (2000) [Tech. Phys. Lett. **26**, 1016 (2000)].
9. S. O. Shiryaeva, Zh. Tekh. Fiz. **71** (2), 27 (2001) [Tech. Phys. **46**, 158 (2001)].
10. S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 163 (2001).
11. S. O. Shiryaeva, Zh. Tekh. Fiz. **72** (4), 15 (2002) [Tech. Phys. **47**, 389 (2002)].
12. A. H. Nayfeh, *Introduction to Perturbation Techniques* (Wiley, New York, 1981; Mir, Moscow, 1984).
13. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 165–171. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 31–37.

GASES AND LIQUIDS

Analytical Solution of the Generalized Smoluchowski Problem Using a New Model Kinetic Equation

A. V. Latyshev and A. A. Yushkanov

Moscow Pedagogical University, ul. Radio 10a, Moscow, 107005 Russia

e-mail: latyshev@orc.ru

Received February 14, 2002

Abstract—A new kinetic equation whose solution yields a correct Prandtl number is suggested. A technique for analytically solving half-space boundary-value problems is demonstrated with the classical Smoluchowski problem of finding temperature and concentration steps. Numerical results indicating the succession of the equation. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION AND DERIVATION OF A NEW MODEL EQUATION

A temperature jump at the gas–condensed phase interface is among the most important phenomena in the kinetic theory of gases. This effect is of crucial importance in many physical processes and should be taken into consideration in determining the thermal conductivity of a gas, studying the dynamics of aerosol particles, etc.

This phenomenon was first studied by Smoluchowski and continues to be a topical problem. There are a number of approaches to studying it, among which those providing analytical solutions are of particular importance. First, analytical approaches were developed by Welander [1]; later, they were extended by Cercignani [2], Sone [3], and others. To date, analytical solutions to the Smoluchowski problem have been found within the Bhatnagar–Gross–Krook (BGK) [4] and ellipsoidal statistical models [5] with a constant collision frequency. Also, an analytical solution to this problem has been obtained within the BGK model with a variable collision frequency (proportional to the molecule velocity) [6].

The Smoluchowski problem borders on the problem of weak evaporation of a gas from a flat surface. Weak evaporation is evaporation with a gas velocity that is much less than the velocity of sound. The latter problem has many applications concerned with heat and mass transfer between a gas and a condensed phase [3]. The two problems differ only in boundary conditions. Therefore, it is convenient to treat them together as the generalized Smoluchowski problem.

The well-known disadvantage of the BGK model is that it gives an unrealistic Prandtl number. For the BGK model with a variable collision frequency, this discrepancy is smaller than for the BGK model where the collision frequency is assumed to be constant. Yet the difference from the realistic Prandtl number still exists.

Note that the Prandtl number is an important parameter characterizing the properties of gases. Therefore, finding an analytical solution to the Smoluchowski problem for the kinetic equation with a variable collision frequency that provides the correct Prandtl number remains on the agenda of researchers.

Let us construct a model kinetic equation with a variable collision frequency that would yield the correct Prandtl number. We will consider only the linearized case, since the Smoluchowski problem is linear in its conventional statement. In the linearized steady-state case, the distribution function can be represented as $f(\mathbf{r}, \mathbf{v}) = f_0(1 + \varphi(\mathbf{r}, \mathbf{v}))$, where f_0 is the Maxwell distribution function, $f_0 = n_s\beta^{3/2}\exp(-\beta v^2)$, $\beta = m/(2kT_s)$, T_s is the surface temperature, and n_s is the concentration of the saturated vapor for evaporation from a surface at a temperatures T_s.

In the general case, the linearized steady-state kinetic equation for φ is given by $\mathbf{C}\nabla\varphi = J[\varphi]$, where $J[\varphi]$ is the linearized collision integral and $\mathbf{C} = \sqrt{\beta}\,\mathbf{v}$ is the dimensionless molecule velocity. For the BGK equation with a constant collision frequency ν, the linearized collision integral has the form

$$J[\varphi] = \sqrt{\beta}\nu\left[2\mathbf{C}\cdot\mathbf{U} + \frac{\delta n}{n_0} + \left(C^2 - \frac{3}{2}\right)\frac{\delta T}{T_0} - \varphi\right].$$

Here,

$$\mathbf{U} = \pi^{-3/2}\int \exp(-C^2)\mathbf{C}\varphi d^3C,$$

$$\frac{\delta n}{n_0} = \pi^{-3/2}\int \exp(-C^2)\varphi d^3C,$$

$$\frac{\delta T}{T_0} = \frac{2}{3}\pi^{-3/2}\int \exp(-C^2)\left(C^2 - \frac{3}{2}\right)\varphi d^3C.$$

For the BGK equation with a collision frequency proportional to the molecule velocity (and with a con-

stant free path $\sim \nu_0^{-1}$), the linearized collision integral is given by [2, 6]

$$J[\varphi] = \sqrt{\beta}\nu_0\left[\frac{3}{4}\sqrt{\pi}C\mathbf{C}\cdot\mathbf{U}_* + \frac{\sqrt{\pi}}{2}C\frac{\delta n_*}{n_0} + \frac{\sqrt{\pi}}{4}C(C^2-2)\frac{\delta T_*}{T_0} - C\varphi\right].$$

Here,

$$\mathbf{U}_* = \pi^{-3/2}\int C'\exp(-C'^2)\mathbf{C}'\varphi d^3C',$$

$$\frac{\delta n_*}{n_0} = \pi^{-3/2}\int C'\exp(-C'^2)\varphi d^3C',$$

$$\frac{\delta T_*}{T_0} = \pi^{-3/2}\int C'\exp(-C'^2)(C'^2-2)\varphi d^3C'.$$

Since all the terms in the expression for the collision integral are proportional to the dimensionless velocity C, the collision frequency of molecules is also proportional to the molecule velocity.

The cases of constant and variable (proportional to the molecule velocity) collision frequencies are particular cases. In general, the collision frequency is a complicated function of the molecule velocity.

Consider a simple nontrivial generalization of the above collision integrals that allows for a more complicated dependence of the collision frequency on the molecule velocity.

Let the linearized collision integral have the form

$$J[\varphi] = \sqrt{\beta}\nu_0[2\alpha_{11}C\mathbf{C}\cdot\mathbf{U}_* + 2\alpha_{12}\mathbf{C}\cdot\mathbf{U}_* + 2\alpha_{21}C\mathbf{C}\cdot\mathbf{U} + 2\alpha_{22}\mathbf{C}\cdot\mathbf{U} + \frac{\sqrt{\pi}}{2}C\frac{\delta n_*}{n_0} + \frac{\pi}{4}C(C^2-2)\frac{\delta T_*}{T_0} - C\varphi\Big]. \quad (1)$$

Here, the right-hand side contains terms that are and are not proportional to the molecule velocity; in other words, the collision frequency is a mixed function of the molecule velocity.

The collision integral must have the following property [7]:

$$\int\exp(-C^2)\varphi J[\psi]d^3C = \int\exp(-C^2)\psi J[\varphi]d^3C, \quad (2)$$

where φ and ψ are arbitrary functions of the molecule velocity.

Property (2) implies the symmetry of the coefficients: $\alpha_{12} = \alpha_{21}$. In addition, collision integral (1) must satisfy the molecule conservation law, momentum conservation law, and energy conservation law [7]. With these requirements satisfied, we come to the set of equations

$$\int\exp(-C^2)J[\varphi]\psi_j(\mathbf{C})d^3C = 0; \quad j = 0, 1, 2, 3, 4, \quad (3)$$

where ψ_j are collision invariants: $\psi_0 = 1$, $(\psi_1, \psi_2, \psi_3) = \mathbf{C}$, and $\psi_4 = C^2$.

System (3) results in a linear relationship between the coefficients α_{ij}. Let $\alpha_{22} = a$ be a free parameter. Then, the other parameters are expressed through a as follows: $\alpha_{12} = \alpha_{21} = -2\alpha a$, $\alpha_{11} = 2\alpha(1 + 2\alpha a)$, and $\alpha = 3\sqrt{\pi}/16$. Also, let $a_1 = -2\alpha a$ and $a_2 = 2\alpha(1 + 2\alpha a)$.

Thus, in terms of the dimensional spatial coordinate introduced by the substitution $x \longrightarrow \nu_0 x$, the steady-state linearized model kinetic equation has the form

$$\frac{\mathbf{C}}{C}(\nabla\varphi) + \varphi(\mathbf{r}, \mathbf{C}) = \frac{1}{2\pi}\int\exp(-C'^2)C'k(\mathbf{C}, \mathbf{C}')\varphi(\mathbf{r}, \mathbf{C}')d^3C' \quad (4)$$

with the kernel

$$k(\mathbf{C}, \mathbf{C}') = 1 + \frac{1}{2}(C^2-2)(C'^2-2) + \frac{4\mathbf{C}\mathbf{C}'}{\sqrt{\pi}CC'}[a + a_1(C + C') + a_2CC'].$$

The parameter a and the Prandtl number are related as

$$\mathrm{Pr} = \frac{8\alpha(1+2a) - 2a}{9\alpha - 2a(1 - 9\alpha^2)}.$$

For $a = 0$, Eq. (4) yields exactly the well-known equation that follows from the BGK model with a velocity-proportional collision frequency [2, 6].

The presence of the parameter a allows for the use of model equation (4) in applications where heat conduction and viscosity are key mechanisms.

Let a gas occupy the half-space $x > 0$ over the gas–condensed phase interface at which the constant-rate evaporation (condensation) and gas–condensed phase heat exchange take place. Away from the interface, a constant temperature gradient K_t normal to it is specified. Let the gradient be small and the rate of evaporation (condensation) be much lower than the velocity of sound; then, the problem can be considered in the linear statement. Outside the Knudsen layer, the gas temperature and concentration profiles have the form $T(x) = T_0 + K_t x$ and $n(x) = n_0 - K_t x$, respectively, and the gas velocity is constant. The parameters to be found are $\varepsilon_t = T_0/T_s - 1$ and $\varepsilon_n = n_0/n_s - 1$ (so-called relative temperature and concentration steps). In the linear approximation, $\varepsilon_t = T_t k_t + T_u U$ and $\varepsilon_n = N_t k_t + N_u U$, where $k_t = K_t/(T\nu_0)$ and $U = \sqrt{\beta}\,u$ are the dimensionless temperature gradient and gas velocity, respectively. We will

seek the dimensionless coefficients T_t, T_u, N_t, and N_u (so-called coefficients of temperature and concentration steps). The statement of this problem is described in detail elsewhere (see, e.g., [1–8]).

For the case of the diffuse reflection of molecules from the walls, the boundary conditions are as follows (the derivation is omitted):

$$\varphi(0, \mu, C) = 0, \quad 0 < \mu < 1,$$
$$\varphi(x, \mu, C) = \varphi_{az}(x, \mu, C) + o(1), \quad x \longrightarrow \infty, \quad -1 < \mu < 0, \tag{5}$$

where

$$\varphi_{az}(x, \mu, C) = \varepsilon_n + \varepsilon_t\left(C^2 - \frac{3}{2}\right) + 2UC\mu + k_t(x - \mu)\left(C^2 - \frac{5}{2}\right) + k_t(b_0 + b_1 C)\mu,$$

$$b_1 = -\frac{1 + 2\alpha a}{8\alpha - 2a(1 - 8\alpha^2)},$$

$$b_0 = \frac{a}{8\alpha - 2a(1 - 8\alpha^2)}, \quad \mu = C_x/C.$$

In this problem, the function φ, boundary conditions (5), and Eq. (4) involve the velocity magnitude C, the cosine μ of the angle between the direction of a molecule and normal to the surface, and one spatial coordinate x. Therefore, Eq. (4) can be integrated over the polar (azimuth) angle from 0 to 2π in the plane (C_y, C_z) to yield

$$\mu\frac{\partial\varphi}{\partial x} + \varphi(x, \mu, C) = \int_{-1}^{1}\int_{0}^{\infty} k(\mu\mu', C, C')\varphi(x, \mu', C')\exp(-C'^2)C'^3 d\mu' dC', \tag{6}$$

$$k(\mu\mu', C, C') = 1 + \frac{1}{2}(C^2 - 2)(C' - 2) + \frac{4\mu\mu'}{\sqrt{\pi}}[aCC' + a_1(C + C') + a_2].$$

Note that the Prandtl number Pr = 2/3 (which is valid for monoatomic gases) correspond to the following values of the parameters:

$$a = \frac{3\alpha}{1 - 6\alpha^2}, \quad a_1 = -\frac{6\alpha^2}{1 - 6\alpha^2}, \quad a_2 = \frac{2\alpha}{1 - 6\alpha^2},$$

$$b_0 = \frac{3}{2}, \quad b_1 = -\frac{1}{2\alpha}.$$

REDUCTION OF THE SMOLUCHOWSKI PROBLEM TO THE ONE-DIMENSIONAL PROBLEM

We will seek a solution to the problem (5)–(6) in the form

$$\varphi(x, \mu, C) = h_1(x, \mu) + Ch_2(x, \mu) + (C^2 - 2)h_3(x, \mu).$$

Then, Eq. (6) is equivalent to the one-dimensional vector equation

$$\mu\frac{\partial h}{\partial x} + h(x, \mu) = \frac{1}{2}\int_{-1}^{1} K_0(\mu, \mu')h(x, \mu')d\mu' \tag{7}$$

for the column vector $h = \mathrm{col}\{h_1, h_2, h_3\}$ with the kernel $K_0(\mu, \mu')$

$$K_0(\mu, \mu') = K_0 + \frac{4}{\sqrt{\pi}}\mu\mu' K_1,$$

$$K_0 = \begin{bmatrix} 1 & 4\alpha & 0 \\ 0 & 0 & 0 \\ 0 & \alpha & 1 \end{bmatrix},$$

$$K_1 = \begin{bmatrix} a + 4\alpha a_1 & 0 & 2\alpha\alpha_1 \\ a_1 + 4\alpha a_2 & 2a_2 + 4\alpha a_1 & 2\alpha a_2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Boundary conditions (5) pass to

$$h(0, \mu) = 0, \quad 0 < \mu < 1,$$
$$h(x, \mu) = h_{as}(x, \mu) + o(1), \quad x \longrightarrow +\infty, \quad -1 < \mu < 0, \tag{8}$$

$$h_{as}(x, \mu) = \begin{bmatrix} \varepsilon_n + \frac{1}{2}\varepsilon_t - \frac{1}{2}k_t(x - \mu) + b_0 k_t\mu \\ b_1 k_t\mu + 2U\mu \\ \varepsilon_t + k_t(x - \mu) \end{bmatrix}.$$

We will simplify the kernel of (7) taking advantage of the mass and energy conservation laws. Eventually, we will arrive at the representation of the kernel of (7) in the form $K(\mu, \mu') = K_0 + \mu\mu' K_2$, where K_0 is the same as before and

$$K_2 = \begin{bmatrix} \frac{4}{\sqrt{\pi}}ac_0 & 0 & 0 \\ 0 & 3c_0(1 + 2\alpha a) & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

with $c_0 = 1 - 9\alpha^2$.

SEPARATION OF VARIABLES AND EIGENSOLUTIONS

Let us separate the variables in Eq. (7). The general Fourier method yields

$$h_\eta(x, \mu) = \exp\left(-\frac{x}{\eta}\right)\Phi(\eta, \mu), \tag{9}$$

where η is a spectral (or separation) parameter, which is generally complex.

Expression (9) immediately leads to the characteristic equation

$$\begin{aligned}(\eta - \mu)\Phi(\eta, \mu) &= \frac{1}{2}\eta K_0 n^{(0)}(\eta) + \frac{1}{2}\mu\eta K_2 n^{(1)}(\eta),\\ n^{(\alpha)}(\eta) &= \int_{-1}^{1}\mu^\alpha\Phi(\eta, \mu)d\mu; \quad \alpha = 0, 1.\end{aligned} \tag{10}$$

Integrating Eq. (10) over μ from –1 to +1 yields $n^{(1)}(\eta) = \eta(E - K_0)n(\eta)$, where E is the unit matrix and $n(\eta)$ is the vector $n^{(0)}(\eta)$. Hence, Eq. (10) can be recast as

$$\begin{aligned}(\eta - \mu)\Phi(\eta, \mu) &= \frac{1}{2}\eta\Delta(\mu\eta)n(\eta),\\ \Delta(x) &= K_0 + xK_3, \quad K_3 = K_2(E - K_2).\end{aligned} \tag{11}$$

Since Eq. (7) is homogeneous, we can put $n(\eta) = \mathrm{col}\{1, 1, 1\}$.

Let $\eta \in (-1, 1)$; then, in the space of generalized functions [9], we find from Eq. (11) the eigenmatrix (eigenfunction) of the characteristic equation

$$\Phi(\eta, \mu) = \frac{1}{2}\eta\Delta(\mu\eta)P\frac{1}{\eta - \mu} + \Lambda(\eta)\delta(\eta - \mu). \tag{12}$$

Here, Px^{-1} is the distribution, i.e., the principal value of the integral upon integrating x^{-1}; $\delta(x)$ is the Dirac function; and

$$\Lambda(z) = E + z\frac{1}{2}\int_{-1}^{1}\frac{\Delta(z\mu)}{\mu - z}d\mu$$

is the dispersion matrix of the problem.

Thus, the continuous spectrum of solutions is found from equality (9), where $\Phi(\eta, \mu)$ is given by relationship (12) for $\eta \in (0, 1)$.

In subsequent analysis, we need the dispersion matrix in explicit form:

$$\Lambda(z) = \begin{bmatrix}\lambda_C(z) & 4\alpha z t(z) - 3c_0 a z^2\lambda_C(z) & 0\\ 0 & \omega(z) & 0\\ 0 & \alpha z t(z) & \lambda_C(z)\end{bmatrix}.$$

Here, $\lambda_C(z) = 1 + zt(z)$ is the Case dispersion function [10], where

$$t(z) = \frac{1}{2}\int_{-1}^{1}\frac{d\tau}{\tau - z}, \quad \omega(z) = 1 + 3c_0(1 + 2\alpha a)z^2\lambda_C(z).$$

By the dispersion function of the problem is meant the determinant of the dispersion matrix $\lambda(z) \equiv \det\Lambda(z) = \lambda_C^2(z)\omega(z)$ (see, e.g., [10]).

The discrete spectrum of Eq. (10) is formed by the zeros of the dispersion function. Let us first find those of $\omega(z)$. Using the principle of argument [11], one can show that the function $\omega(z)$ has two simple real zeros $\pm\eta_0$ lying beyond the cut (–1, 1) near its extremes. Let $\eta = \eta_0$; then, from Eq. (11), it follows

$$\Phi(\eta_0, \mu) = \frac{1}{2}\eta_0\Delta(\mu\eta_0)\frac{1}{\eta_0 - \mu}n(\eta_0). \tag{13}$$

Substituting the function

$$h_{\eta_0}(x, \mu) = \exp\left(-\frac{x}{\eta_0}\right)\Phi(\eta_0, \mu), \tag{14}$$

where $\Phi(\eta_0, \mu)$ is given by (13), into Eq. (7), we find that the function $h_{\eta_0}(x, \mu)$ is a solution of (7) if and only if $n(\eta_0)$ satisfies the homogeneous solution

$$\Lambda(\eta_0)n(\eta_0) = \Theta, \quad \Theta = \mathrm{col}\{0, 0, 0\}. \tag{15}$$

Equation (15) has a nonzero solution because $\lambda(\eta_0) = \det\Lambda(\eta_0) = 0$.

Let us split Eq. (12) into three scalar ones to find the vector $n(\eta_0)$:

$$n(\eta_0) = \begin{bmatrix}-4\alpha\eta_0 t(\eta_0) + 3c_0 a\eta_0^2\lambda_C(\eta_0)\\ \lambda_C(\eta_0)\\ -\alpha\eta_0 t(\eta_0)\end{bmatrix}. \tag{16}$$

Thus, the eigensolution of the discrete spectrum that corresponds to the zero η_0 is constructed. It is given by equalities (13), (14), and (16).

The dispersion function $\lambda(z)$ has one more zero, which is the quadruple zero $\eta_i = \infty$. To this zero, there correspond four linearly independent discrete solutions, which are column vectors in the expansion of $h_{as}(x, \mu)$.

SOLUTION TO THE BOUNDARY-VALUE PROBLEM: REDUCTION TO A VECTOR BOUNDARY-VALUE PROBLEM

A solution to the problem stated by (7) and (8) will be sought as the expansion in eigensolutions of the dis-

crete and continuous spectra:

$$h(x,\mu) = A_0 h_{\eta_0}(x,\mu) + h_{as}(x,\mu) + \int_0^1 \exp\left(-\frac{x}{\eta}\right)\Phi(\eta,\mu)A(\eta)d\eta, \quad -1<\mu<1, \quad x>0 \tag{17}$$

or (in a more detailed form)

$$h(x,\mu) = A_0 h_{\eta_0}(x,\mu) + h_{as}(x,\mu) + \frac{1}{2}\int_0^1 \exp\left(-\frac{x}{\eta}\right)\frac{\eta\Delta(\mu\eta)}{\eta-\mu}A(\eta)d\eta + \exp\left(-\frac{x}{\mu}\right)\Lambda(\mu)A(\mu)\Theta_+(\mu).$$

Here, A_0, ε_n, and ε_t are unknown constants (the coefficients of the discrete spectrum); $A(\eta)$ is an unknown vector function (the coefficient of the continuous spectrum); and $\Theta_+(\mu) = 1$ (for $0<\mu<1$) or 0 (for $-1<\mu<0$).

Substituting expansion (17) into the first boundary condition in (8) yields a singular integral equation with the Cauchy kernel in the interval (0, 1):

$$A_0 h_{\eta_0}(0,\mu) + h_{as}(0,\mu) + \frac{1}{2}\int_0^1 \frac{\eta\Delta(\mu\eta)}{\eta-\mu}A(\eta)d\eta + \Lambda(\mu)A(\mu) = \Theta. \tag{18}$$

Let us introduce the auxiliary vector function

$$N(z) = \frac{1}{2}\int_0^1 \eta\Delta(\eta z)A(\eta)\frac{d\eta}{\eta-z}, \tag{19}$$

for which, according to the Sokhotsky formulas,

$$N^+(\mu) - N^-(\mu) = \pi i\mu\Delta(\mu^2)A(\mu), \quad 0<\mu<1. \tag{20}$$

Using the boundary values of $N(z)$ and $\Lambda(z)$, we reduce Eq. (18) to the Riemann–Hilbert vector boundary-value problem

$$B^+(\mu)[N^+(\mu) + A_0 h_{\eta_0}(0,\mu) + h_{as}(0,\mu)] = B^-(\mu)[N^-(\mu) + A_0 h_{\eta_0}(0,\mu) + h_{as}(0,\mu)], \quad 0<\mu<1. \tag{21}$$

Consider the related homogeneous boundary-value problem

$$B^+(\mu)X^+(\mu) = B^-(\mu)X^-(\mu), \quad 0<\mu<1. \tag{22}$$

Here, $X(z)$ is an unknown matrix that is analytical in the plane with a cut along the segment [0, 1]. Taking into account that the dispersion matrix can be represented in the form

$$\Lambda(z) = \lambda_C(z)\Delta(z^2) + K_4, \quad K_4 = \begin{bmatrix} 0 & -4\alpha & 0 \\ 0 & 1 & 0 \\ 0 & -\alpha & 0 \end{bmatrix},$$

we find that

$$B(z) = \lambda_C E + \frac{1}{3c_0(1+2\alpha a)z^2}K_4.$$

Omitting the diagonalization of problem (22) (see [6]), we write the matrix

$$X(z) = \begin{bmatrix} U(z) & 4\alpha(U(z)-V(z)) & 0 \\ 0 & V(z) & 0 \\ 0 & \alpha(U(z)-V(z)) & U(z) \end{bmatrix}$$

without proof. Here,

$$U(z) = z\exp(-u(z)), \quad u(z) = \frac{1}{\pi}\int_0^1 \frac{\Theta(\tau)}{\tau-z}d\tau,$$

$$\Theta(\tau) = -\frac{\pi}{2} - \arctan\frac{2\lambda_C(\tau)}{\pi\tau},$$

$$V(z) = z\exp(-v(z)), \quad v(z) = \frac{1}{\pi}\int_0^1 \frac{\varepsilon(\tau)}{\tau-z}d\tau,$$

$$\varepsilon(\tau) = -\frac{\pi}{2} - \arctan\frac{2\omega(\tau)}{3c_0(1+2\alpha a)\pi\tau^3}.$$

Now we return to the solution of inhomogeneous problem (21). With the homogeneous problem, we transform it into the problem of finding an analytical vector function from its zero step at a cut,

$$[X^+(\mu)]^{-1}[N^+(\mu) + A_0 h_{\eta_0}(0,\mu) + h_{as}(0,\mu)] = [X^-(\mu)]^{-1}[N^-(\mu) + A_0 h_{\eta_0}(0,\mu) + h_{as}(0,\mu)], \quad 0<\mu<1.$$

Taking into consideration the behavior of the matrices and vectors involved in this problem at the point η_0 and at the point $\eta_i = \infty$, we arrive at its general solution:

$$N(z) = -A_0 h_{\eta_0}(0,z) - h_{as}(0,z) + X(z)\left[\frac{B}{z-\eta_0} + C\right]. \tag{23}$$

Here, the vectors B and C have the arbitrary components B_i and C_i, respectively (i = 1, 2, and 3). For solution (23) to be taken as auxiliary function (19), we will eliminate the singularities of the general solution.

SOLVABILITY CONDITIONS AND NUMERICAL CALCULATIONS

To remove the pole at the point η_0 in Eq. (23), we require that

$$\frac{1}{2}A_0\eta_0\Delta(\eta_0^2)n(\eta_0) + X(\eta_0)B = \Theta.$$

This equation is equivalent to three scalar equations, from which we found the vector B:

$$B = \frac{A_0\eta_0}{2V(\eta_0)}\begin{bmatrix}-4\alpha\\ 1\\ -\alpha\end{bmatrix}.$$

Let us expand the matrix function $X(z)$ in the vicinity of the point $\eta_i = \infty$:

$$X(z) = zE - \begin{bmatrix}U_1 & 4\alpha(U_1 - V_1) & 0\\ 0 & V_1 & 0\\ 0 & \alpha(U_1 - V_1) & U_1\end{bmatrix} + o(1),$$
$$|z| \longrightarrow \infty.$$

Here,

$$U_1 = -\frac{1}{\pi}\int_0^1 \Theta(\tau)d\tau, \quad V_1 = -\frac{1}{\pi}\int_0^1 \varepsilon(\tau)d\tau.$$

Now, it is easy to expand $N(z)$ in the vicinity of the point $\eta_i = \infty$:

$$N(z) = zN_1 + N_0 + o(1), \quad |z| \longrightarrow \infty.$$

Here,

$$N_1 = \begin{bmatrix}-k_t(1/2 + b_1) + C_1\\ -(2U + b_1k_t) + C_2\\ k_t + C_3\end{bmatrix},$$

$$N_0 = \begin{bmatrix}-\varepsilon_n - \frac{1}{2}\varepsilon_t - \frac{1}{2}A_0 3c_0 a\eta_0^2 n_1(\eta_0)\\ + B_1 - U_1C_1 - 4\alpha(U_1 - V_1)C_2\\ \frac{1}{2}A_0 3c_0(1 + 2\alpha a)\eta_0^2 n_2(\eta_0) + B_2 - V_1C_2\\ -\varepsilon_t + B_3 - \alpha(U_1 - V_1)C_2 - U_1C_3\end{bmatrix}.$$

To remove the pole of $N(z)$ at the point $\eta_i = \infty$, we construct the vector C:

$$C = \begin{bmatrix}k_t(1/2 + b_0)\\ 2U + b_1k_t\\ -k_t\end{bmatrix}.$$

Now we find the auxiliary function $N(z)$ at the point $\eta_i = \infty$:

$$N(\infty) = \frac{3c_0}{2}\int_0^1 \eta^2\begin{bmatrix}aA(\eta)\\ (1 + 2\alpha a)A_2(\eta)\\ 0\end{bmatrix}d\eta.$$

From the condition $N(\infty) = N_0$, we arrive at the set of equations

$$\frac{3c_0}{2}a\int_0^1 \eta^2 A(\eta)d\eta = -\varepsilon_n - \frac{1}{2}\varepsilon_t - \frac{1}{2}A_0 3c_0 a\eta_0^2 n(\eta_0) + B_1 - U_1C_1 - 4\alpha(U_1 - V_1)C_2, \quad (24)$$

$$-\frac{3c_0}{2}(1 + 2\alpha a)\int_0^1 \eta^2 A_2(\eta)d\eta = -\frac{1}{2}A_0 + B_2 - V_2C_2, \quad (25)$$

$$\varepsilon_t = B_3 - \alpha(U_1 - V_1)C_2 - U_1C_3. \quad (26)$$

Substituting general solution (23) into Sokhotsky formula (20), we find the integrand in (24) and (25):

$$\frac{3c_0}{2}(1 + 2\alpha a)\eta^2 A_2(\eta)d\eta = \frac{1}{2\pi i\eta}[V^+(\eta) - V^-(\eta)]\left(\frac{B_2}{\eta - \eta_0} + C_2\right). \quad (27)$$

Let

$$J(\eta_0) = \frac{1}{2\pi i}\int_0^1 [V^+(\eta) - V^-(\eta)]\frac{d\eta}{\eta - \eta_0}.$$

Applying the contour integration procedure, one can show that

$$J(\eta_0) = V(\eta_0) - \eta_0 + V_1.$$

Integrating (27) in view of (25) and (26) yields

$$\varepsilon_n = 2U\left(\frac{7}{2}\beta - \frac{a(V_1 - \eta_0)}{1 + 2\alpha a}\right) + k_t\left(-U_1 + \frac{7}{2}b_1\beta + \frac{a(V_1 - \eta_0 - U_1)}{8\alpha - 2a(1 - 9\alpha^2)}\right), \quad (28)$$

$$\varepsilon_t = 2U\beta + k_t(U_1 + b_1\beta), \quad (29)$$

where $\beta = \alpha(V_1 - \eta_0 - U_1)$.

With $a = 0$, formulas (28) and (29) give

$$\varepsilon_t = 2U\beta + k_t\left(U_1 - \frac{2\beta}{3\sqrt{\pi}}\right),$$

$$\varepsilon_n = 2U\left(\frac{7}{2}\beta\right) + k_t\left(-U_1 - \frac{7\beta}{3\sqrt{\pi}}\right),$$

in complete agreement with well-known formulas (6.1) in [6].

Let a in (28) and (29) correspond to Pr = 2/3. Then, in a form convenient for calculation, we have

$$\varepsilon_t = 2U\alpha(V_1 - \eta_0 - U_1) + k_t\frac{1}{2}(3U_1 - V_1 + \eta_0),$$
$$\varepsilon_n = U\alpha(V_1 - \eta_0 - 7U_1) + k_t\frac{1}{4}(-3U_1 - V_1 + \eta_0). \quad (30)$$

Numerical calculations show that $\eta_0 = 1 + \varepsilon$, where $\varepsilon = 1.3 \times 10^{-17}$; $U_1 = 0.71045$; and $V_1 = 0.99326$. Hence, from (30),

$$\varepsilon_t = 1.06904k_t - 0.23835(2U),$$
$$\varepsilon_n = -0.53115k_t - 0.82749(2U).$$

Now we pass to the dimension variables, defining the free path of molecules as $l = (\eta/\rho)\sqrt{\pi m/2kT_0}$, where η is the kinematic viscosity and ρ is the density of a gas [8]. Then, the computational formulas for temperature and concentration steps take the form

$$\varepsilon_t = C_t l K_t + 2\sqrt{\beta}C_t^u u, \quad \varepsilon_n = C_n l K_t + 2\sqrt{\beta}C_n^u u.$$

Here,

$$C_t = \frac{15}{8}(1.06904) = 2.00445, \quad C_t^u = -0.23835,$$
$$C_n = \frac{15}{8}(-0.53115) = -0.99591,$$
$$C_n^u = -0.82749.$$

For comparison, results [6] for gases where the collision frequency is proportional to the molecule velocity (i.e., the free path of molecules is constant) are as follows: $C_t = 1.99885$, $C_n = -0.99657$, $C_t^u = -0.23687$, and $C_n^u = -0.82905$.

Results [12] for gases with a constant molecular collision frequency are $C_t = 2.20262$, $C_n = -1.23035$, $C_t^u = -0.22436$, and $C_n^u = -0.84350$.

Finally, we give the results $C_t = 2.1133$, which was obtained by numerically solving the complete Boltzmann equation [13] within the hard sphere model, and $C_t = 2.20576$ [14], which was obtained by using the 13-moment kinetic model with a constant molecular collision frequency. Note that the results cited are given in terms of the molecule free path at Pr = 2/3, which is adopted in this paper.

CONCLUSION

The numerical results obtained in this paper for temperature jumps lie between those found within the BGK model with a constant collision frequency and those found within the BGK model with a constant free path. Thus, our model can be considered as a model with an intermediate character of molecular collisions. Its important advantage is the use of the correct Prandtl number. Otherwise, calculated results must be converted in view of the discrepancy between the true and obtained Prandtl numbers [8]. Such a procedure is to some extent uncertain. Results obtained with our model, which handles the correct Prandtl number, are free of this drawback.

Note also that the accuracy of our model is close to that of the 13-moment kinetic model, which is today very popular.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (project no. 99-01-00336).

REFERENCES

1. P. Welander, Ark. Fys. **7**, 507 (1954).
2. C. Cercignani, Transp. Theory Stat. Phys. **6** (1), 29 (1977).
3. Y. Sone, J. Phys. Soc. Jpn. **45**, 315 (1978).
4. A. V. Latyshev, Prikl. Mat. Mekh. **54**, 581 (1990).
5. A. V. Latyshev and A. A. Yushkanov, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 1, 163 (1992).
6. A. V. Latyshev and A. A. Yushkanov, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 140 (1996).
7. A. V. Latyshev and A. A. Yushkanov, Pis'ma Zh. Tekh. Fiz. **26** (23), 16 (2000) [Tech. Phys. Lett. **26**, 1034 (2000)].
8. C. Cercignani, *Theory and Application of the Boltzmann Equation* (Elsevier, New York, 1975; Mir, Moscow, 1978).
9. V. S. Vladimirov, *Generalized Functions in Mathematical Physics* (Nauka, Moscow, 1976).
10. K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967; Mir, Moscow, 1972).
11. F. D. Gakhov, *Boundary Value Problems* (Nauka, Moscow, 1977; Addison-Wesley, Reading, Mass., 1966).
12. A. V. Latyshev and A. A. Yushkanov, in *Mathematical Models of Non-Linear Excitations, Transfer, Dynamics, and Control in Condensed Systems and Other Media*, Ed. by L. A. Uvarova, A. E. Arinstein, and A. V. Latyshev (Kluwer Academic, New York, 1999).
13. S. K. Loyalka, Transp. Theory Stat. Phys. **20**, 237 (1991).
14. T. Soga, Phys. Fluids **29**, 976 (1986).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 172–176. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 38–42.

GASES AND LIQUIDS

Determination of Hydrodynamic Drag from a Trajectory of Inertial Ship Motion

A. S. Devyatisilny, V. M. Dorozhko, and V. M. Grinyak

Institute of Automation and Control Processes, Far East Division, Russian Academy of Sciences, Vladivostok, 690041 Russia

e-mail: devyatis@iacp.dvo.ru

Received April 24, 2002

Abstract—A model for estimating hydrodynamic drag from a ship's inertial travel and/or velocity is suggested. Full-scale experimental data for sea-going ships are discussed. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The development of methods for describing moving object–environment interaction is of undeniable current interest [1, 2].

The aim of this work is to estimate the hydrodynamic drag from full-scale experiments. This problem is of great importance in designing ship's hull lines and traffic control systems, in estimating the efficiency of object exploitation, etc.

The towing [3, 4] and inertial [1] methods of estimating the hydrodynamic drag on objects are most widely used in full-scale experiments. The former can be applied only in specially equipped testing grounds. In contrast, the inertial method does not require sophisticated facilities and is compatible with current on-line surveillance systems. Therefore, the inertial method can be viewed as the basis for an approach to estimating the hydrodynamic drag from an inertial motion trajectory and, hence, deserves careful examination.

STATEMENT OF THE PROBLEM AND MODEL CONCEPTS

First, let us note the disadvantages typical of the towing method and the advantages of the method suggested in this paper. In the towing method for measuring the hydrodynamic drag, (i) it is necessary to use a tug and towing facilities, (ii) the hydrodynamic drag vs. velocity dependence is derived by repeat towings with various velocities, (iii) the sea-forming properties of a ship are estimated by additionally measuring its seas, and (iv) there is basically no way of excluding the effect of the tug's seas on drag measurements.

While being free of these disadvantages, the inertial method in the form stated in [1] makes it possible to estimate only the sea component of the hydrodynamic drag.

In view of the aforesaid and taking into account that the inertial method is easier to apply, we formulate requirements for a modified inertial method that would allow the derivation of the hydrodynamic drag vs. ship velocity dependence. These are the following: (i) The hydrodynamic drag should be estimated from a single inertial motion trajectory; (ii) since the acceleration of object inertial motion is generally very low (≈ 0.02 m/s^2), trajectory data processing techniques must provide a sufficient accuracy; (iii) a modified inertial method must allow for the identification of motion transients; and (iv) a modified inertial method must allow one to compare the sea-forming properties of objects.

Formally, the problem is reduced to the estimation of the unmeasurable component (acceleration) from the measurable components of the motion vector (position and/or velocity) and then to finding the drag from the known acceleration.

It follows that, in terms of the starting theoretical mechanical concepts, our problem can obviously be treated as the first problem of mechanics and, thus, is the classical inverse problem of dynamics for an object moving in a medium.

We are now coming to the formal statement of the problem. The hydrodynamic drag r is conventionally considered as a function of the steady-state velocity: $r = r_x(v)$ [3, 4]. We will adhere to this approach, but the corresponding function is derived by converting the acceleration of inertial motion to the hydrodynamic drag with the well-known relationship [3, 4]

$$r_x = -(1 + n_r)\ddot{x} \quad \text{or} \quad r_x = -(1 + n_r)\dot{v}, \qquad (1)$$

where r_x is the specific hydrodynamic drag per unit object mass (hereafter drag), x is the linear coordinate along which the object moves (1D motion alone is considered), and n_r is a correction factor taking into account the unsteadiness of the motion and object geometry.

The rigorous theoretical–mechanical interpretation of the factor n_r is straightforward and well known [3, 4].

1063-7842/03/4802-0172$24.00 © 2003 MAIK "Nauka/Interperiodica"

However, with the applied aspect of the problem borne in mind, it is appropriate to represent it in the empirical form

$$n_r = b_0 + b_1 v + \ldots + b_\phi v^\phi + \ldots + b_\gamma v^\gamma, \qquad (2)$$

where $b_\phi = b_{\phi,0} + b_{\phi,1}\psi + b_{\phi,2}\psi^2 + \ldots + b_{\phi,s}\psi^\theta$ are coefficients related to the object geometry, namely, to its length L and volume displacement V_{vol}; hence, $\psi = L/V_{\text{vol}}^{1/3}$.

In this work, the coefficient n_r in the form of (2) is given for the case $\gamma = \theta = 2$ (see below) as the result of the analysis of typical relationships [4] that generalize full-scale test data for a seagoing cargo ship.

As follows from (1) and (2), the drag $r_x(t)$ and, hence, $r_x(v)$ (according to the approach adopted) can be estimated if the position $x(t)$ and velocity $v(t)$ or acceleration $\ddot{x}(t)$ (or $\dot{v}(t)$) are known. Consider the usual practical case when the motion of an object is characterized by measuring $v(t)$ and/or $x(t)$. Assuming that the measurements are performed at discrete time instants t_i^* ($i = \overline{1, n}$) within an interval $[t_1^*, t_n^*]$ where the initial time t_1^* corresponds to the onset of the inertial motion of the object and introducing the normalized time $t \in [-1, 1]$ ($t = (t^* - t_1^* - S_m)/S_m$, where $S_m = (t_n^* - t_1^*)/2$), we represent the measurement models as

$$x(t_i) = \sum_{j=0}^{k-1} h_{xj}(t_i) p_{x,j+1}, \quad k < n, \qquad (3)$$

$$v(t_i) = \sum_{j=0}^{m-1} h_{vj}(t_i) p_{v,j+1}, \quad m < n, \qquad (4)$$

where $\varepsilon_x(t_i) = \varepsilon_{xi}$ and $\varepsilon_v(t_i) = \varepsilon_{vi}$ are random Gaussian instrumental measurement errors, $\varepsilon_x : N(0, \sigma_x^2)$ and $\varepsilon_v : N(0, \sigma_v^2)$, and $h_j(t)$ are the basis functions ($j = 1, 2, \ldots,$).

Expressions (3) and (4) can be written in the vector matrix form

$$Z_x = H_x p_x + \varepsilon_x, \qquad (5)$$

$$Z_v = H_v p_v + \varepsilon_v, \qquad (6)$$

where the meaning and components of the vectors and matrices are quite obvious.

Equations (5) and (6) can be viewed as declarative models of two problems, one of which appeals to the measurement of only the object position on the trajectory and the other, to the measurement of only the velocity.

The appeal to the simultaneous measurement of the position and velocity (which may be necessary to improve the accuracy of the specific drag estimate) leads to a model of the third problem:

$$\begin{pmatrix} Z_x \\ Z_v \end{pmatrix} = \begin{pmatrix} H_x \\ \dot{H}_x \end{pmatrix} p_{xv} + \begin{pmatrix} \varepsilon_x \\ \varepsilon_v \end{pmatrix}, \qquad (7)$$

or, in terms of the obvious designations, $Z_{xv} = H_{xv} p_{xv} + \varepsilon_{xv}$.

The Hadamard requirements for the correction of the statements [5] for these three problems are easily satisfied if the bases of the spaces of definition x and v are represented by sets of the functions $h_j(t)$ that are orthogonal within the segment [–1, 1]. In this case, solutions to problems (5) and (6) are perfectly conditioned and highly accurate. As such functions, we will use Hayes orthogonal polynomials [6].

Note that the dimension of the linear space of solutions to problems (5)–(7), k and m, depends on the complexity of object motion (or on the complexity of velocity variation in our case, since we are considering 1D motion in the physical space) and is not known *a priori*. Therefore, when finding the final results, we should carry out an additional procedure, namely, to check the hypothesis G : $\dim(p_x) = \dim(p_{xv}) = k$ or $\dim(p_v) = m$.

Thus, the formal statement of the starting problem, that is, the use of the inertial method to estimate the hydrodynamic drag as a function of velocity, basically implies the solution of one of the problems (5), (6), or (7) subject to the two alternative hypotheses for motion complexity and the need to find the domain of definition for this function in view of the actual object motion.

FORMAL SOLUTION AND RESULT SUITABLE FOR APPLICATION

In the general form, a model of a solution to the above problem can be represented by the following tuple:

$$\{\hat{x}, \hat{\dot{x}}, \hat{\ddot{x}}, \hat{v}, \hat{\dot{v}}, \hat{k}, \hat{\sigma}_x, \hat{m}, \hat{\sigma}_v, \hat{r}_x(\hat{v}), \sigma_r(\hat{v}), [\hat{v}_{\min}, \hat{v}_{\max}]\}, \qquad (8)$$

where the cap means the estimate of the corresponding quantity, σ_r is the rms estimation error of the drag $\hat{r}_x$, and $[\hat{v}_{\min}, \hat{v}_{\max}]$ is the interval of definition of the function $\hat{r}_x(\hat{v})$ with regard to actual object motion.

Obviously, the first nine components $\{\hat{x}, \hat{\dot{x}}, \hat{\ddot{x}}, \hat{v}, \hat{\dot{v}}, \hat{k}, \hat{\sigma}_x, \hat{m}, \hat{\sigma}_v\}$ of model (8), which refer to a solution to problem (5), (6), or (7), play a decisive role, since the others are obtained by straightforward calculations.

Consider first problem (5) or (6). Solving any of them by the least-squares method [5], we arrive at the

following estimates (the subscripts x and v indicating the type of measurement are omitted):

$$\hat{p}_{j+1} = \left(\sum_i w_i h_j(t_i) Z_i\right) \Big/ \sum_i w_i h_j^2(t_i);$$
$$\hat{\sigma}_p^2 = \hat{\sigma}^2 \left(\sum_i w_i h_j^2(t_i)\right)^{-1}, \quad (9)$$

where $i = \overline{1, n}$ and $j = \overline{0, k-1}$ for position measurements or $j = \overline{0, m-1}$ for velocity measurements, Z_i is the value of an ith component of the vector Z, $w_i \equiv 1$ according to the previously adopted hypothesis that measurements of the same type are of the same accuracy, $\hat{\sigma}^2$ is the estimated variance of the instrumental error, and $\hat{\sigma}_p^2$ is the estimated variance of the error in determining the vector p.

Now we pass to problem (7). Note first of all that a set of polynomials that are derived from the initial set of orthogonal polynomials by time differentiation is generally no longer orthogonal. With this taken into account, a solution to problem (7) can be represented in the matrix form

$$\hat{p}_{xv} = (H_{xv}^T Q_{xv}^{-1} H_{xv})^{-1} H_{xv}^T Q_{xv}^{-1} Z_{xv},$$
$$Dp_{xv} = \hat{\sigma}^2 (H_{xv}^T Q_{xv}^{-1} H_{xv})^{-1}, \quad (10)$$

where Dp_{xv} is the variance matrix for the vector $\hat{p}_{xv}$ and $Q_{xv} = M[\varepsilon_{xv}\varepsilon_{xv}^T]$.

One should bear in mind that velocity measurements combined with position measurements improve the estimates of all the components of the spectra p_x and p_v in object motion models (5) and (6) except for the first component.

As the value of $\hat{\sigma}^2$, we take the value of $\hat{\sigma}_x^2$ or $\hat{\sigma}_v^2$ subject to the hypothesis G : $\dim(p) = \xi$ (ξ takes the value k for $\hat{\sigma}_x^2$ or m for $\hat{\sigma}_v^2$). These values are calculated from the expression (the subscripts x, v, and xv are omitted)

$$\hat{\sigma}^2 = \|Z - H\hat{p}\|^2/(n - \xi). \quad (11)$$

The hypothesis is checked in order to find the minimal value of ξ in the range of steady-state values of $\hat{\sigma}^2$ by comparing values of $\hat{\sigma}^2$ adjacent in ξ with each other or with an *a priori* value of $\hat{\sigma}_a^2$ provided that it is known with a desired degree of confidence.

In view of (1), (2), (4)–(7), (9), and (10), as well as the relationship $dt = S_m^{-1} dt^*$, the estimate of the specific drag $\hat{r}_x(\hat{v})$ as a function of the estimated velocity in problems (5)–(7) can be found by eliminating t_i^* from sets of expressions which follow:

$$\hat{r}_x(t_i^*) = \ddot{W}_x \hat{p}_x / S_m^2,$$
$$\hat{r}_x(t_i^*) = \dot{W}_v \hat{p}_v / S_m,$$
$$\hat{r}_x(t_i^*) = \ddot{W}_x \hat{p}_{xv} / S_m^2,$$
$$\hat{v}(t_i^*) = \dot{H}_x \hat{p}_x / S_m;$$
$$\hat{v}(t_i^*) = H_v \hat{p}_v;$$
$$\hat{v}(t_i^*) = \dot{H}_x \hat{p}_{xv} / S_m. \quad (12)$$

The quality of the estimates of the function $\hat{r}_x(\hat{v})$ in problems (5)–(7) is specified by the diagonal elements $\sigma_r^2(\hat{v})$ of the variance matrices $D_{rx}(\hat{v})$, which are obtained by eliminating t_i^* from the sets of expressions

$$D_{rx}(t_i^*) = \ddot{W}_x D_{px}^* \ddot{W}_x^T,$$
$$D_{rx}(t_i^*) = \dot{W}_v D_{pv} \dot{W}_v^T,$$
$$D_{rx}(t_i^*) = \ddot{W}_x D_{pxv}^* \ddot{W}_x^T,$$
$$\hat{v}(t_i^*) = \dot{H}_x \hat{p}_x / S_m;$$
$$\hat{v}(t_i^*) = H_v \hat{p}_v;$$
$$\hat{v}(t_i^*) = \dot{H}_x \hat{p}_{xv} / S_m, \quad (13)$$

where (with the subscripts x, v, and xv omitted) $W = -(1 + n_r)H$, $D_p = \hat{\sigma}^2 (H^T H)^{-1}/S_m^2$, and $D_p^* = D_p/S_m^2$.

EXPERIMENTAL RESULTS AND DISCUSSION

The model that estimates the specific drag as a function of steady-state velocity values was experimentally verified based on acceptance test data for the Vitus Bearing cargo ship.

The fixed parameters of the model were the ship's length along the waterline $L = 150$ m, volume displacement $V_{vol} = 20200$ m^3, initial ship velocity $v_1 = 8.2$ m/s, and inertial motion duration 1500 s. The variable parameters of the motion, the ship's position and velocity under calm conditions, were measured with standard measuring means (radar and log). The maximal number of measurements was $n = 500$. The components of the vector b for $\gamma = \theta = 2$ were obtained by approximating typical velocity dependences of the coefficient n_r [3, 4] with regard to the ship's length and volume displacement: $b_0 = 0.2017 - 0.0321\psi + 0.0010\psi^2$, $b_1 = (-0.4871 + 0.1193\psi - 0.0034\psi^2)/\sqrt{gL}$, and $b_2 = (5.9608 - 1.3184\psi + 0.0679\psi^2)/(gL)$, where $\psi = L/V_{vol}^{1/3}$

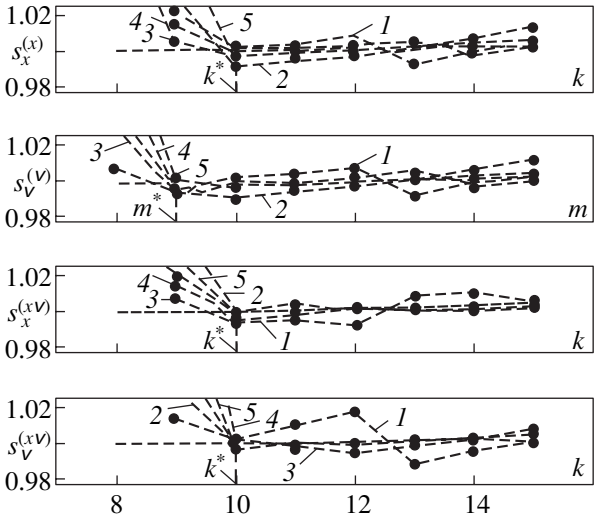


Fig. 1. Normalized estimated variances $s_\nu^{(\zeta)}$ of measurement errors for n = (*1*) 100, (*2*) 200, (*3*) 300, (*4*) 400, and (*5*) 500 readings. The subscript $\nu = \{x, v\}$ indicates the quantity under consideration: position or velocity, respectively; the superscript $\zeta = \{x, v, xv\}$ indicates the type of measurement: position measurements, velocity measurements, or both.

and $g = 9.81$ m/s^2. The maximum permissible value of the rms deviation $\sigma_{r,\max}$ normalized to the maximal value of the function $\hat{r}_x$ was taken to be $w_{\max} = \sigma_{r,\max}/\hat{r}_{x,\max} = 0.01$.

The dimension of the vectors p for various estimation schemes was defined as the minimal value of k^* or m^* in the range where the parameter given by (11) takes steady-state values (Fig. 1): $s = \hat{\sigma}^2/\tilde{\sigma}^2$, where $\tilde{\sigma}^2$ is the mean value of the estimated variance in the range of its steady-state values.

To analyze the behavior of the drag calculated from the measured parameters of the inertial motion, we also present the specific drag normalized to its maximal value for another vessel, a Rowdy yacht. In this case, the drag was evaluated from the time profiles of the velocity [1] for two starting velocities: five and six knots (curves c_2 and c_3, respectively, in Fig. 2).

It is known [4] that the dependence of the total drag on the steady-state ship velocity can be described by the relationship $R_x = C(0, 5\rho S v^2)$, where C is the dimensionless coefficient of drag, ρ is the density of the liquid, and S is the wetted surface area of the ship. Note that $C = C_{\text{vis}} + C_s$, where C_{vis} and C_s are the coefficients of the viscous and sea components of the total drag, respectively. Taking into account the monotonic run of the C_{vis} vs. v curve [3], as well as the constancy of ρ and S, we can conclude that the variation of both the total and specific drags is defined by the coefficient C_s, which generally grows at Froude numbers $\text{Fr} = v/\sqrt{gL} \le 0.2$ [3, 4, 7]. Therefore, to see how the sea-forming effect of the inertial motion affects the specific drag, we will consider its dependence on the relative velocity (Froude number).

Figure 2 shows the estimated specific drags (given by (12)) on the inertial motion of the Vitus Bearing ship (curve c_1, $n = 100$) and Rowdy yacht (curves c_2, c_3) normalized to their maxima. Also, the derivatives of the drags with respect to the Froude number are shown (curves c_1'–c_3'). The maximal values of the derivatives correspond to the inflection points b_1–b_3 in the curves c_1–c_3. In view of the above-mentioned behavior of the coefficient C_s, it follows from curve c_1 that the range of the drag for the Vitus Bearing ship correlates with the relative velocity interval $[\text{Fr}_{\min}, \text{Fr}_{\max}]$, where $\text{Fr}_{\max} =$

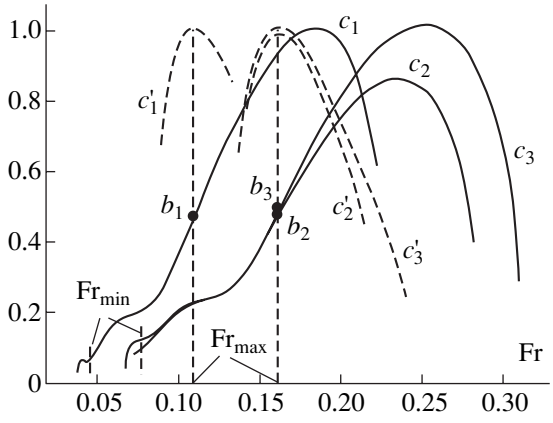


Fig. 2. Normalized estimated specific drags and their derivatives with respect to Froude number. Curve c_2 is normalized to the maximal value of curve c_3.

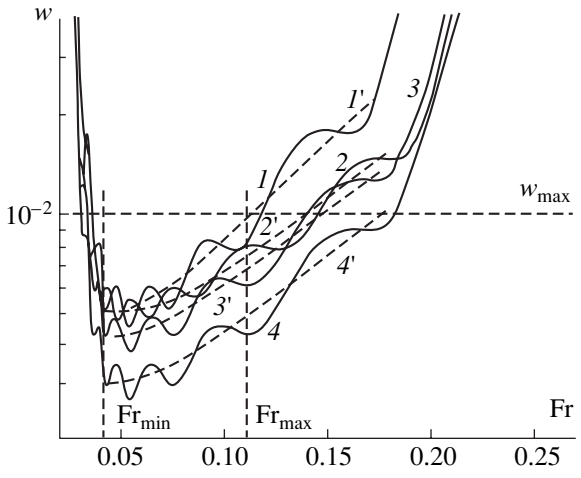


Fig. 3. Normalized rms deviations w of the estimated specific drags determined from (*1*) position measurements; (*2*) velocity measurements; and (*3*, *4*) combined measurements for the number of measurements n and $n/2$, respectively. Curves *1'*–*4'* are the averaged curves *1*–*4*.

$\arg\{\max_{\mathrm{Fr}}(\partial\hat{r}_x/\partial\mathrm{Fr})\}$. It is interesting that the position of the points b_2 and b_3 does not depend on the starting velocities; that is, $\mathrm{Fr}_{\max}$ is a measure of the sea-forming properties of ships. The choice of $\mathrm{Fr}_{\min}$ depends on our knowledge of factors affecting the ship's motion and, hence, drag at low velocities (instability of longitudinal motion, natural environmental fluctuations, etc.).

Figure 3 demonstrates the Fr dependences of the normalized rms deviations (given by (13)) $w = \sigma_r/\hat{r}_{x,\max} < w_{\max}$ for the Vitus Bearing ship ($n = 100$). The curves indicate the high accuracy of the specific drag estimates due to the use of the least-squares regression of observations on an orthogonal basis. The maximal accuracy of the estimates (the least variance, curves *3* and *4*) for the specific drag is achieved when position and velocity measurements are combined.

CONCLUSION

Thus, as follows from our experimental investigation, the inertial method developed in this paper, unlike the towing method, makes it possible to derive the velocity dependence of the drag and compare the sea-forming properties of ships of different type by processing data for one trajectory of their inertial motion. Obviously, a Rowdy yacht has a better design of hull lines and exhibits less pronounced sea-forming properties, since the effect of its seas on the specific drag disappears at higher Froude numbers than in the case of the Vitus Bearing ship. The maximum of the derivative of the specific drag indicates the termination of a transient due to inertial motion. Below this point, the dependences studied are free of any disturbances. In the case of the towing method, conversely, foreign effects on measurements (for example, the tug's seas) are basically impossible to exclude.

Clearly, results obtained by our method can be used for the rapid estimation of specific drag actual values because of its simplicity. If the estimates are made at more or less regular intervals, they can serve as a means for inspecting the condition (fouling) of the ship's body (in view of the fact that the viscous component of the total drag prevails at small Froude numbers).

REFERENCES

1. J. N. Selness, Mar. Technol. **26** (3), 210 (1989).
2. A. B. Podlaskin, Zh. Tekh. Fiz. **68** (6), 32 (1998) [Tech. Phys. **43**, 644 (1998)].
3. *Handbook on the Ship Theory*, Ed. by Ya. I. Voĭtkunskiĭ (Sudostroenie, Leningrad, 1985), Vol. 1.
4. Ya. I. Voĭtkunskiĭ, *Resistance to Ship's Motion* (Sudostroenie, Leningrad, 1988).
5. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Englewood Cliffs, N.J., Prentice Hall, 1974; Nauka, Moscow, 1989).
6. G. A. F. Seber, *Linear Regression Analysis* (Wiley, New York, 1977; Mir, Moscow, 1980).
7. A. Sh. Gotman, *Determination of Wave Resistance and Optimization of Ship's Hull Lines* (Novosib. Gos. Akad. Vodn. Transp., Novosibirsk, 1995), Part 1.

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 177–184. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 43–50.

GASES AND LIQUIDS

Local Effect of Electric and Magnetic Fields on the Position of an Attached Shock in a Supersonic Diffuser

S. V. Bobashev*, R. V. Vasil'eva*, A. V. Erofeev*, T. A. Lapushkina*, S. A. Poniaev*, and D. M. Van Wie**

** Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*
e-mail: Serguei.Poniaev@mail.ioffe.ru
*** John Hopkins University, Laurel, Maryland, 20723 USA*

Received July 30, 2002

Abstract—Effective ways for controlling shock wave configurations by means of external actions are sought. One such way is a local effect of electric and magnetic fields. In this paper, the local effect of external fields is implemented by current localization in a limited region of a diffuser. The experiment is carried out in a diffuser providing the complete internal compression of the gas with a Mach number at the inlet M = 4.3. As a working medium, a xenon plasma is used. The plasma flow is formed in a shock tube equipped with an accelerating nozzle. Two ways of current localization are tested. In the first one, the diffuser inlet is a short channel of Faraday generator type. In this case, the ponderomotive force basically decelerates or accelerates the flow depending on the direction of the electric field. In the second way, the current flows through a narrow near-wall region between adjacent electrodes. In this case, the ponderomotive force compresses or expands the gas. In both cases, it is shown that the angle of an attached shock due to MHD interaction can be both decreased and increased. The central problem with the MHD control of shock waves is near-electrode and near-wall phenomena. © *2003 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

At present, various approaches to controlling new-generation hypersonic flying vehicles being designed are being widely discussed [1]. In this work, we experimentally study the possibility of controlling shocks at the inlet of a supersonic diffuser with a magnetohydrodynamic (MHD) technique. This problem was initiated by the AJAX concept [1, 2] and developed in a number of theoretical [3–8] and experimental studies [9–13]. This paper is a continuation of this series of experimental investigations. *I–V* characteristics described in [13] showed a great potential drop near the electrodes, which was compensated for by applying an external electric field. Therefore, in the experiment considered below, the flow pattern is formed both as a result of MHD interaction and by applying an electric field.

When investigating the influence of external fields applied to various diffuser sections [12], we found that the efficiency of field application is higher if the field is applied to the diffuser inlet. By efficiency, we mean the ratio of a certain change in the attached shock angle to the power consumption. Therefore, it is natural to continue the investigations in this direction and to study different ways of local action. In our new experiment, an external effect is localized by confining the electric current in a limited region of the diffuser. We tested two ways of current localization. In the one case, the current passed through oppositely placed electrodes; in the other, through electrodes placed on the same wall of the MHD channel. The different arrangement of the electrodes corresponded to various ways of force action.

EXPERIMENT

The experimental setup is schematically shown in Fig. 1. It consists of a shock tube with high-pressure

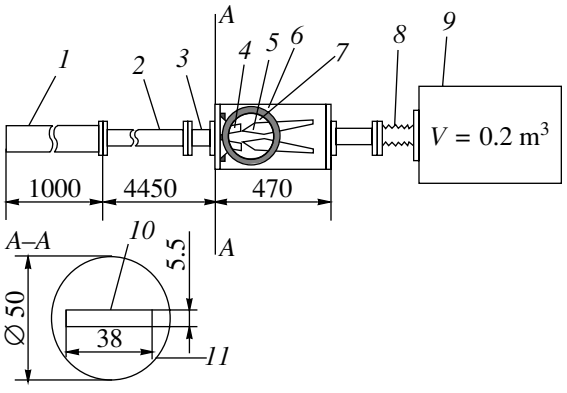


Fig. 1. Experimental setup. (*1*) High-pressure chamber, (*2*) low-pressure chamber, (*3*) test section of the low-pressure chamber, (*4*) plane nozzle, (*5*) diffuser, (*6*) vacuum chamber, (*7*) optical window, (*8*) bellows, (*9*) gas receiver, (*10*) nozzle inlet cross section, and (*11*) shock tube cross section. Dimensions are given in millimeters.

1063-7842/03/4802-0177$24.00 © 2003 MAIK "Nauka/Interperiodica"

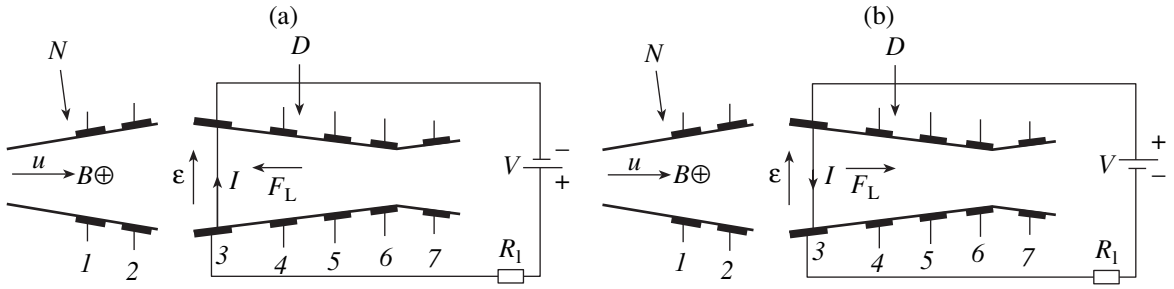


Fig. 2. First way of current localization: (a) decelerating regime and (b) accelerating regime. *N*, nozzle; *D*, diffuser. The figures are electrode numbers, and the arrows indicate the directions of the current *I*, emf ε, ponderomotive force F_L, and flow *u*.

chamber *1* designed for a hydrogen pressure of up to 150 atm and low-pressure chamber *2* ≈50 mm in diameter and 4.5 m in length filled with a heavy inert gas to a pressure of 20–50 torr. The chambers are separated by a metallic diaphragm. The low-pressure chamber is terminated by test section *3* with a piezoelectric pressure transducer and windows to extract the radiation of the shock-heated gas. At the end of the test section, there is the throat of plane nozzle *4*, which is separated from the test section by a thin lavsan diaphragm. The gas flow accelerated in the nozzle enters diffuser *5* with electrodes on its opposite walls. The nozzle and the diffuser are placed in vacuum chamber *6*. Schlieren patterns of the flow are recorded through optical windows *7*. The gasdynamic path extends beyond the vacuum chamber through bellows *8* to gas receiver *9*. The setup is described in detail in [9].

As the shock-heated gas becomes stagnant at the nozzle end face, its temperature increases and thermal ionization occurs. The ionized gas passes through narrow nozzle throat *10* at face end *11* of the test portion (cross section *A–A* in Fig. 1) and enters divergent nozzle *4*, where the gas is accelerated to specified Mach numbers. The expanding plasma cools down, but the gas ionization in the nozzle lowers rather slowly because of the slow recombination in inert gas plasmas [14]; therefore, the plasma conductivity appropriate for MHD experiments is provided. The supersonic flows of relaxing inert-gas plasmas and their interaction with a magnetic field are detailed elsewhere [15].

A pulsed magnetic field of induction 1.5 T directed transversely to the flow was produced by the discharge of a 0.0125-F capacitor bank through coaxial Helmholtz coils 300 mm in diameter placed on the vacuum chamber walls. The discharge duration is 3.5 ms, and the magnetic field induction can be considered constant over a period of 500 μs. The latter time interval was used to synchronize the process. An electric pulse of an appropriate amplitude and duration was produced by the "long line" discharge. To visualize the flow, we designed an optical Schlieren system. The filming of the process was performed with a VSK-5 high-speed camera with a Podmoshenskiĭ illuminant [16]. The exposure time per frame was 2.7 μs. A higher time resolution (30 ns) was achieved by using an OGM-20 ruby laser as a light source and taking single pictures.

The experiment was carried out in xenon at a Mach number of the shock wave front in the shock tube of *8* and an initial xenon pressure in the low-pressure chamber of 30 torr. The design gas parameter at the diffuser inlet were M = 4.3, the xenon density $\rho = 0.127$ kg/m^3, the flow velocity $u = 1.55$ km/s, and the plasma conductivity $\sigma = 600$ S/m. The plasma efflux duration was approximately equal to 500 μs.

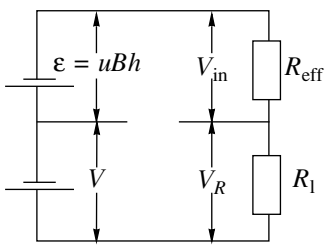


Fig. 3. Equivalent electric circuit.

FIRST WAY OF CURRENT LOCALIZATION

In the case considered, the effect of magnetic and electric fields on the plasma flow is localized in a short section at the diffuser inlet [18], as shown in Fig. 2. The current flows only through the third pair of the electrodes at the diffuser inlet. In this case, the diffuser inlet operates as a Faraday MHD generator and an electromotive force (emf) $\varepsilon = uBh$ is induced at the inlet (here, *u* is the flow velocity, *B* is the magnetic induction, and *h* is the electrode spacing). A voltage *V* of either polarity can be applied to the electrodes from an external source. When the emf induced by the magnetic field and the external electric field are codirected (Fig. 2a), we deal with the so-called decelerating regime; when the emf and the field are opposing (Fig. 2b), the regime is conventionally referred to as accelerating. The equivalent electric circuit of the setup is shown in Fig. 3. Its basic feature is that the source of external voltage *V* is connected in series to the magnetically induced emf. The sum of the emf and applied voltage partially drop across load resistor R_l and partially, across the internal

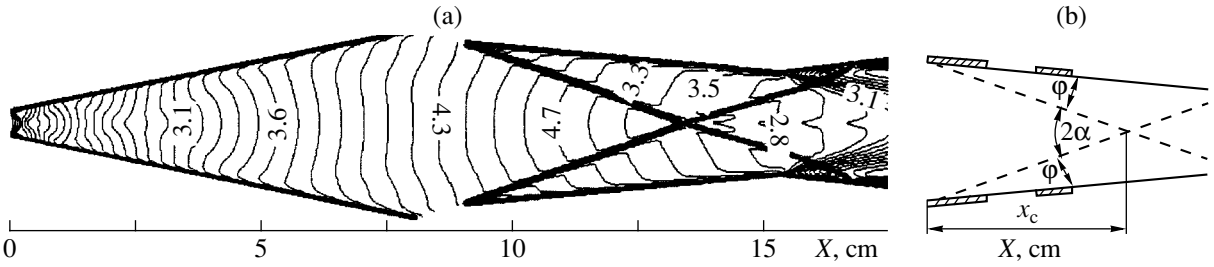


Fig. 4. Shock wave configuration in the diffuser ($B = 0$ and $V = 0$). (a) Calculated distribution of the flow Mach number over the nozzle and diffuser [3] and (b) basic characteristics of the flow pattern in a supersonic diffuser. 2α is the angle between interacting attached shocks, φ is the angle between an attached shock and the diffuser wall, and x_c is the point of crossing the attached shocks.

resistance R_{eff} of the interelectrode gap. The total of these two resistances is close to the long-line wave impedance, from which the voltage V is picked up. In this case, Ohm's law has the form

$$IR_{\text{eff}} = (1-k)(uBh \pm V), \tag{1}$$

where R_{eff} is the interelectrode gap resistance and k is the load factor given by

$$k = \frac{R_l}{R_l + R_{\text{eff}}}. \tag{2}$$

The plus sign before the second term in Eq. (1) corresponds to MHD deceleration; the minus sign, to MHD acceleration.

Consider factors affecting the flow exposed to magnetic and electric fields and try to evaluate their relative effects on the flow parameters, especially on the flow Mach number, because its value is mainly responsible for the shock angle.

The MHD effect has two elementary components. The first one is the work of the mass force with a density juB, which decelerates or accelerates the flow. The other is the removal (supply) of the electric energy produced by the magnetically induced emf. Its density per unit volume equals $kjuB$. The external electric field heats the plasma via the release of Joule heat with a density $Q = (1 - \text{k})jV/h$. Our qualitative analysis of the problem will rely on the one-dimensional solution [17]. In [17], all elementary effects on the flow are represented in the form of additive terms with coefficients depending on the flow parameters. For example, the equation for a Mach number M of the flow exposed to a magnetic field of strength uB and an electric field of strength V/h can be written in the form

$$(\text{M}^2 - 1)\frac{d\ln \text{M}^2}{dx} = \frac{j}{a^2 \rho u}\left(f_{\text{M}}(\text{M})uB + f_{\text{E}}(\text{M})\frac{V}{h}\right), \tag{3}$$

where j is the current density, a is the speed of sound, ρ is the gas density, and u is the flow velocity.

The coefficients characterizing the MHD effect (f_{M}) and the influence of the external electric field (f_{E}) are defined as follows:

$$\begin{aligned} f_{\text{M}}(\text{M}) &= \mp\gamma(2 + (\gamma-1)\text{M}^2) \pm (\gamma-1)(1+\gamma\text{M}^2)k, \\ f_{\text{E}}(\text{M}) &= -(\gamma-1)(1+\gamma\text{M}^2)(k-1), \end{aligned} \tag{4}$$

where $\gamma = C_p/C_v$ is the ratio of specific heats at constant pressure and constant volume.

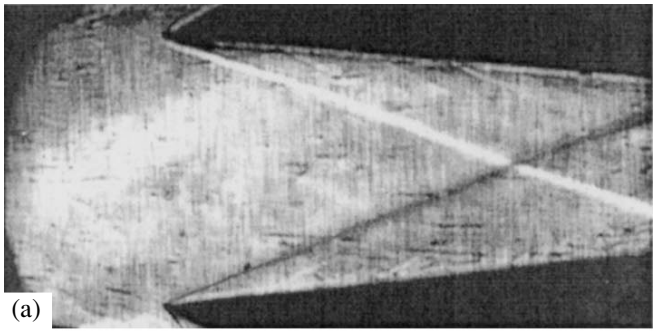
In Eq. (4) the upper minus before the first term and the upper plus before the second one correspond to the decelerating regime, while the inverse means acceleration.

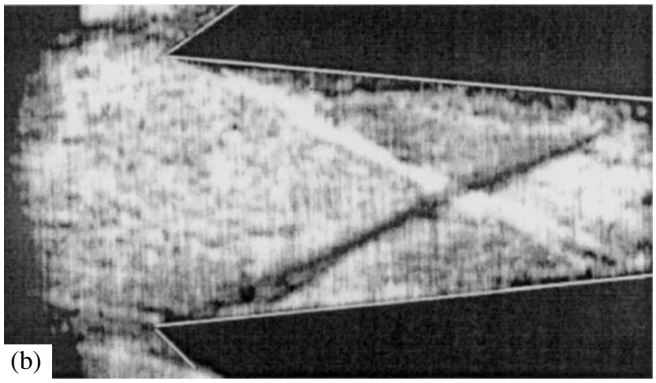
From Eq. (3) it follows that the effect of the magnetic field dominates that of the electric field if the magnetically induced emf ($\varepsilon = uBh$) and the external voltage meet the inequality

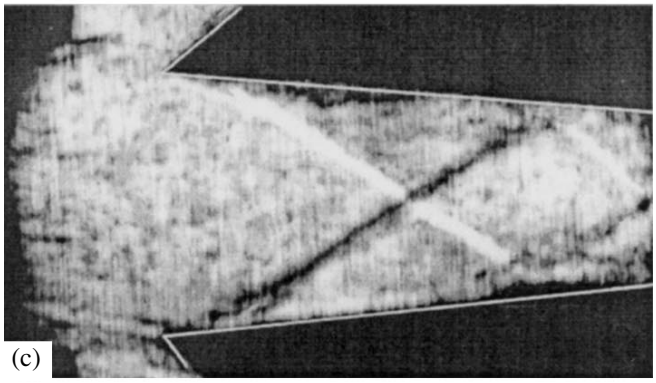
$$\begin{aligned} \frac{uBh}{V} &> \frac{f_{\text{E}}(\text{M})}{f_{\text{M}}(\text{M})} \\ &= \frac{(\gamma-1)(1-k)(1+\gamma\text{M}^2)}{\gamma(2+(\gamma-1)\text{M}^2) - k(\gamma-1)(1+\gamma\text{M}^2)}. \end{aligned} \tag{5}$$

Equation (5) shows that at M > 5 the Mach number is governed largely by the MHD effect rather than by the electric field for $uBh/V > 1$ and any load factor. At M = 1 and $\gamma = 5/3$, the effect of the magnetic field prevails when $uBh/V > 2(1-k)/(5-2k)$. Thus, the relative role of MHD interaction increases as the flow Mach number decreases. The greater the load factor, the stronger the MHD effect. Let us estimate the limiting values of V/uBh corresponding to our experiment by using formula (5). In the experiment $k \approx 0.2$ and the Mach numbers calculated at the inlet and outlet are equal to 4.3 and 3, respectively. Then, the limiting values of V/ε turn out to be equal to 1.15 and 1.3. Thus, the interval separating the regions of prevalence of the magnetic and electric field effects is $1.3 > V/\varepsilon > 1.15$.

Parameters of the shock wave configuration that characterize the response of the flow to the external effect are clarified from the shock pattern in the diffuser (Fig. 4). The diffuser is formed by two wedge-shaped walls with an angle of 5.5°. An attached shock makes


(a)


(b)


(c)

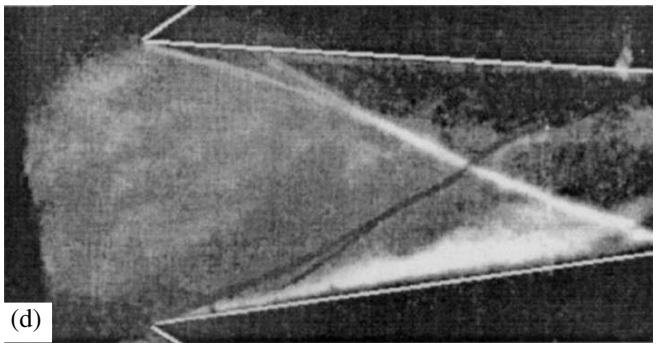

(d)

Fig. 5. Examples of shock wave configurations in the flow subjected to electric and magnetic fields. (a) $V = 0$ and $B = 0$, (b) $V = 120$ V and $B = 0$, (c) $V = 125$ V and $B = -1.3$ T (decelerating regime), (d) $V = 150$ V and $B = -1.3$ T (accelerating regime).

an angle φ with the diffuser wall. Two shocks attached to the opposite diffuser walls cross one another at a distance x_c from the diffuser inlet and reflect from one another at an angle 2α. Each of the three parameters φ, x_c, and α can serve as an indicator of an external effect. For example, when the flow is decelerated and the flow Mach number decreases, the shock angle grows, the point of shock intersection approaches the diffuser inlet, and the angle between the shocks increases. When the flow is accelerated, these parameters vary in the opposite direction.

Shlieren pictures of the steady-state flows are shown in Fig. 5. Figure 5a displays the shock wave configuration without the fields ($V = 0$ and $B = 0$). In this case, $\varphi = 15°$, $x_c = 43$ mm, and $\alpha = 21.5°$. The situation when the magnetic field is absent ($B = 0$) and the electric field is applied ($V = 120$ V) is demonstrated in Fig. 5b. It is seen that the increase in the flow Mach number due to heating is accompanied by an increase in the shock angle φ, decrease in the distance x_c, and increase in the angle α between the shocks. Figure 5c shows the Schlieren picture of the flow exposed to both magnetic and electric fields ($B = -1.3$ T and $V = 125$ V) (decelerating regime). Here, the joint action of the fields causes stronger flow deceleration; that is, the angle α increases and the distance x_c decreases. Figure 5d shows the Schlieren pattern of the flow in the accelerating regime at $V = -150$ V and $B = -1.3$ T. It is seen that additional shocks practically merge with the attached shocks probably due to the strong current-related heating of the near-wall layer. As a whole, the angle α increases and the distance x_c decreases as compared with the case $V = 0$ and $B = 0$; that is, the flow does not accelerate. It seems that the flow deceleration due to the Joule heating of the gas is stronger than the flow acceleration caused by the ponderomotive force in this case. However, as follows from Fig. 4b, the joint action of the accelerating MHD effect and electric field weakens the decelerating effect caused by the Joule heating of the gas.

In a series of experiments, we maintained the magnetic induction B at –1.3 T and applied voltages of various amplitude and polarity to the electrodes. With such a magnetic field, the magnetically induced emf at the diffuser inlet was $\varepsilon = 65$ V. The variation of the attached shock position with electric and magnetic field is shown in Fig. 6 as the dependences of the half-angle between the shocks and of the current on the relative voltage V/ε.

Let us turn to the I–V characteristics presented in Fig. 6. The solid line C is the I–V characteristic calculated on the assumption that the plasma resistance is equal to the value specified at the diffuser inlet. This dependence demonstrates that the ideal I–V characteristic is linear. However, the experimental dependence $I(V/\varepsilon)$ is essentially nonlinear and the current, almost without exception, is considerably less than that predicted theoretically. All this is evidence of the important role of near-electrode processes in current passage.

Moreover, it turns out that this role is not the same at $V > 0$ and $V < 0$. From the I–V characteristic, one can estimate the near-electrode potential drop ΔV at low currents [13]. It is determined from the initial portion of the I–V characteristic up to the value of V at which the derivative of the dependence $I(V)$ starts to decrease. At $V > 0$ (the decelerating regime), $\Delta V/\varepsilon \approx 1.1$ (see Fig. 6); that is, the near-electrode potential drop is close to or slightly greater than the magnetically induced emf. The same also follows from the fact that the current through the interelectrode gap is practically absent when $V = 0$ and the plasma is affected by the emf alone. The absence of the current can be explained by the formation of unneutralized space charge near the cold nonemitting electrodes. The emission of electrons from a cold cathode is possible [18] when microarcs and cathodic spots appear. In the fast processes, this can occur at a current density of about 50 A/cm^2. In our experiments, such a current density is achievable only at relatively high voltages applied to the electrodes. Besides, cold gasdynamic boundary layers and possible separation of the flow may prevent current passage.

It is interesting that the reversal of the sign of the applied voltage modifies the dependence $I(V/\varepsilon)$. It should also be noted that at $|V/\varepsilon| > 1$, the current in the range of negative V grows faster than in the positive range. The differential plasma resistance determined from the I–V characteristic for $V/\varepsilon < -1$ turned out to be roughly 0.25 Ω, while for $V/\varepsilon > 1$, it grows to 0.43 Ω. Thus, the I–V characteristic suggests that near-wall effects in the decelerating regime are more pronounced than in the accelerating one.

The vertical lines in Fig. 6 show the values of V/ε separating regions with different dominant effects. There are three such demarcation lines, which separate regions I–IV.

First, let us consider these boundaries. The first boundary covers the interval $1.15 < V/\varepsilon < 1.3$, the second is the line $V/\varepsilon = -1$, and the third is the line $V/\varepsilon = -1.15$. In region I ($V/\varepsilon > 1.3$), the relationships between the applied voltage and emf are such that the flow deceleration due to Joule heating by the external electric field is stronger than the deceleration caused by MHD interaction. In region II ($-1.0 < V/\varepsilon < 1.15$), the ponderomotive force is the basic reason for the flow deceleration. In region III (narrow range $-1.15 < V/\varepsilon < -1.0$), the relationships between the external voltage and magnetically induced emf are such that the ponderomotive force can accelerate the flow. In region IV ($V/\varepsilon < -1.15$), the flow deceleration due to Joule heating by the external electric field dominates the accelerating action of the magnetic field.

Consider how the shock wave at the inlet responds to the variations of the voltage across the electrodes. First, let us analyze the behavior of the angle α as a function of the applied voltage for $V > 0$ (Fig. 6). In region II, the angle α varies insignificantly. This is because the current in this region is insufficient to considerably decelerate the flow owing to near-electrode processes. The larger increase in the angle α is observed in region I, where the current is higher and the flow deceleration is stronger because of the gas heating by the external electric field and the action of the ponderomotive force.

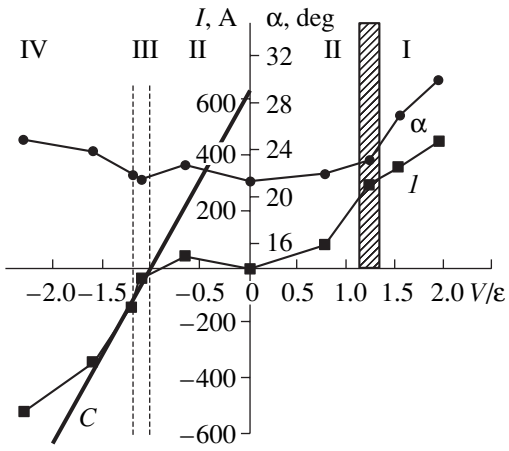


Fig. 6. Electric current and angle α as functions of the relative voltage V/E (ε = 65 V). I, region with the dominating role of the electric field in flow deceleration; II, region with the dominating role of the MHD effect in flow deceleration; III, region with the MHD acceleration of the flow; and IV, region with the dominating role of the electric field in the accelerating MHD regime.

Now we trace the variation of the angle α for $V < 0$ in Fig. 6. Here, the behavior of α in region II is noteworthy: α increases noticeably at I = 40 A. This seems to be rather unexpected because at $V > 0$ in the same region, α remains practically unchanged. The reason is that during the MHD interaction, the current oscillated, while at other V, the current remained constant. Since there was no exact synchronization between process recording and current passage, it is not improbable that the flow pattern was recorded at a current exceeding its average value. Therefore, we observed the flow deceleration under the dominant effect of the ponderomotive force. In region III, the current changes direction, but the flow deceleration, that is, the decrease in the angle α as compared with its value at $V = 0$, is absent. In region IV, the effect of the external fields becomes noticeable, as demonstrated by the increase in α. The increase in α indicates that, as was expected, the flow deceleration in region IV takes place, because the gas heating by the external field dominates the accelerating action of the ponderomotive force. Thus, in our experiment, Joule heating in the external electric field has the most noticeable effect on the flow both in the decelerating and accelerating regimes.

The experimental data allow us to separate the effects of flow acceleration and deceleration caused by the ponderomotive force from the decelerating effect due to Joule heating in an external electric field. This is done as follows. We select three regimes with the same

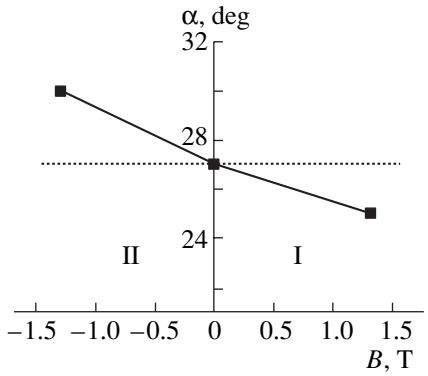


Fig. 7. Variation of the angle between the attached shocks that is caused by magnetic induction reversal. The dotted line indicates the angle α when $B = 0$ and the current is the same as in the presence of the magnetic field ($B = \pm 1.3$ T). I, accelerating regime; II, decelerating regime.

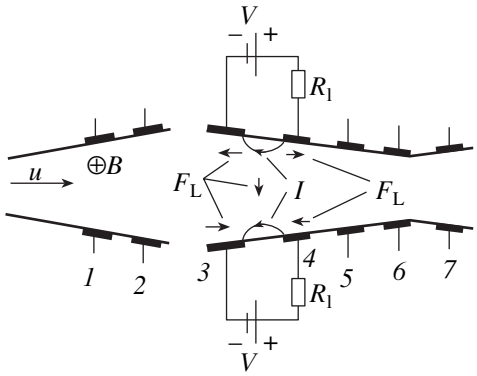


Fig. 8. Second way of current localization. The arrows indicate the directions of the current I, magnetic field B, Lorentz force F_L, and flow velocity u.

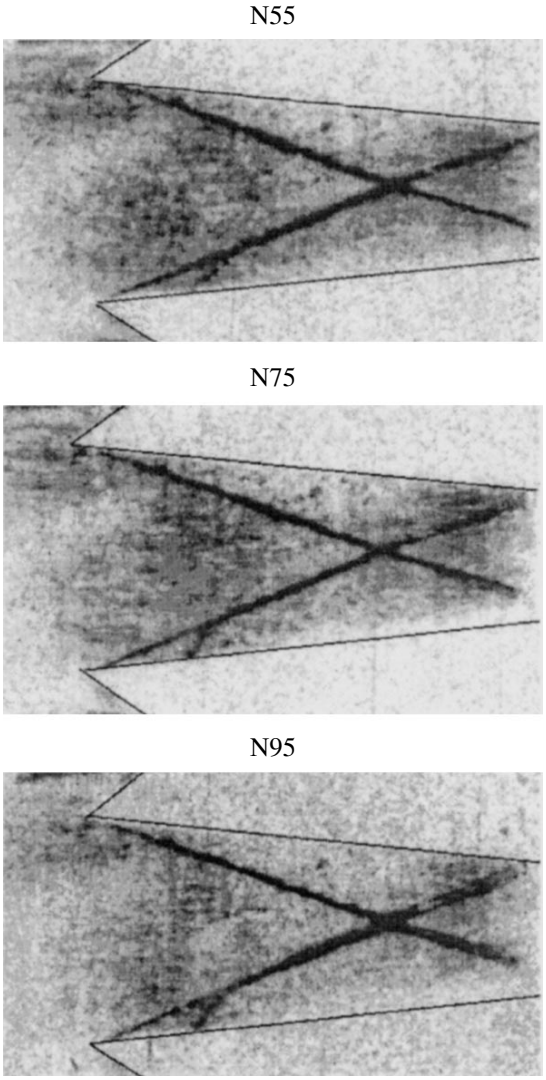


Fig. 9. Steady-state flow pictures. The figures above are picture numbers. The exposure time is 1.9 μs, the time interval between the pictures is 5.7 μs.

current $I = (500 \pm 50)$ A and two oppositely directed and one zero ($B = 0$) ponderomotive force. The ponderomotive force causing the flow deceleration corresponds to a negative value of the magnetic induction ($B = -1.3$ T); the accelerating ponderomotive force, to the positive magnetic induction $B = 1.3$ T. The results of such an approach are presented in Fig. 7, where the angle α is plotted against the magnetic induction. The horizontal dotted line depicts the value $\alpha = 27°$ (corresponding to $B = 0$). The angle α corresponding to the decelerating regime is larger and that corresponding to the accelerating regime is smaller than the value at $B = 0$, as it should be.

To conclude, we see no basic obstacles in controlling shocks at the inlet with the ponderomotive force other than near-wall effects.

SECOND WAY OF CURRENT LOCALIZATION

The scheme of the experiment is shown in Fig. 8. The current passes through a local region behind the attached shocks near the diffuser walls between electrodes *3* and *4*. In Fig. 8, the directions of the current and Lorentz force near the upper and lower diffuser walls are shown by arrows. The Lorentz force near the opposite walls is seen to act in opposite directions. At the upper wall, the Lorentz force tends to decrease the gas pressure, while at the lower wall, it compresses the gas. In both cases, the gas also warms up owing to the Joule heat. The current passing through the upper and lower electrodes is equal to 400 and 500 A, respectively. The magnetic induction is $B = 1.3$ T.

Figure 9 displays a number of steady-state flow pictures. The asymmetric arrangement of the attached

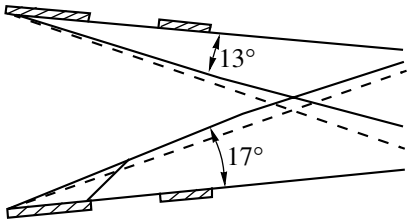


Fig. 10. Comparison of the attached shock positions. Solid lines correspond to the Schlieren patterns in Fig. 9; dashed lines show the shock positions at $V = 0$ and $B = 0$.

shocks is noteworthy. To clarify changes in the shock wave structure, Fig. 10 shows the shocks in the absence of external effects (dashed lines) and those depicted in Fig. 9 (solid lines). At $B = 0$ and $V = 0$, the shocks are symmetric about the flow axis and make an angle $\varphi = 15°$ with the diffuser walls. After the magnetic field has been applied and the current from an external source has passed through local region *3–4*, the shock angle near the upper wall decreases to 13°; that is, the shock shifts nearer to the wall. At the lower wall, the shock angle increases to 17°; i.e., the shock moves away from the wall. Such a behavior of the near-wall shocks is explained by the fact that the directed action of the Lorentz force near the upper wall decreases the gas pressure and the shock approaches the wall. Near the lower wall, the Lorentz force acts in the opposite direction, raising the pressure, and the shock moves away from the wall.

CONCLUSIONS

It is shown that the attached shock angle in a diffuser can be controlled by locally affecting the flow with external magnetic and electric fields.

In the first way of control, the effect is localized at the diffuser inlet, since the current passes only through the electrodes located at the inlet (Faraday scheme of electrode connection). The ranges of the applied voltage and magnetically induced emf are found where the flow deceleration (acceleration) is governed by MHD interaction or Joule heating in an external electric field. The experimental results suggest that the one-dimensional theory provides good estimates of the domains where the MHD interaction or the electric field dominates.

The decelerating and accelerating effects of the ponderomotive force and the decelerating action of an external electric field are separated. The attached shock angle can be increased or decreased by appropriately applying the ponderomotive force.

With the Faraday scheme of electrode connection, the role of near-wall effects in the decelerating regime is greater than in the accelerating one.

In the second way of control, the current flows through adjacent electrodes downstream from the attached shocks and the Lorentz force is localized mainly in a narrow near-wall region. In this case, the Lorentz force either compresses or expands the gas. In the experiment, both an increase and decrease in the attached shock angle was observed.

Comparing the effects of external fields on the location of attached shocks in the two ways of current localization, one can note that, in spite of a lesser change in the shock angle with the second way (at approximately the same power consumption), it seems to be preferable, since in this case the flow core is disturbed less and, consequently, one may expect total pressure losses to be lower.

ACKNOWLEDGMENTS

The work is financially supported by the EOARD.

REFERENCES

1. E. P. Gurijanov and P. T. Harsha, in *Proceedings of 7th Aerospace Planes and Hypersonic Technology Conference, Norfolk, 1996* (AIAA Pap. **96-4609**, 1996).
2. V. L. Fraĭshtad, A. L. Kuranov, and E. G. Sheĭkin, Zh. Tekh. Fiz. **68** (11), 43 (1998) [Tech. Phys. **43**, 1309 (1998)].
3. Yu. P. Golovachev, A. A. Schmidt, and S. Yu. Suschikh, in *Proceedings of 2nd Workshop on Magneto-Plasma-Aerodynamics in Aerospace Applications* (Inst. Vys. Temp. Ross. Akad. Nauk, Moscow, 2000).
4. A. Vatazhin, V. Kopchenov, and O. Goushkov, in *Proceedings of 2nd Workshop on Magneto-Plasma-Aerodynamics in Aerospace Applications* (Inst. Vys. Temp. Ross. Akad. Nauk, Moscow, 2000), p. 64.
5. D. I. Brichkin, A. L. Kuranov, and E. G. Sheikin, in *Proceedings of 8th AIAA International Space Planes and Hypersonic Systems and Technologies Conference, Norfolk, 1998* (AIAA Pap. **98-1642**, 1998).
6. Yu. P. Golovachev and S. Yu. Sushikh, Zh. Tekh. Fiz. **70** (2), 28 (2000) [Tech. Phys. **45**, 168 (2000)].
7. S. O. Macheret, M. N. Shneider, and R. B. Miles, in *Proceedings of 39th Aerospace Sciences Meeting and Exhibition, Reno, 2001* (AIAA Pap. **2001-0492**, 2001).
8. V. A. Bityurin, J. T. Lineberry, V. G. Potebnia, *et al.*, in *Proceedings of 28th AIAA Plasmadynamics and Lasers Conference, Atlanta, 1997* (AIAA Paper **97-2393**, 1997).
9. S. V. Bobashev, E. A. D'yakonova, A. V. Erofeev, *et al.*, AIAA Pap. **2000-2647** (2000).
10. S. V. Bobashev, R. V. Vasil'eva, E. A. D'yakonova, *et al.*, Pis'ma Zh. Tekh. Fiz. **27** (2), 63 (2001) [Tech. Phys. Lett. **27**, 71 (2001)].
11. S. V. Bobashev, A. V. Erofeev, T. A. Lapushkina, *et al.*, in *Proceedings of 3rd Workshop on Magneto-Plasma-Aerodynamics in Aerospace Applications* (Inst. Vys. Temp. Ross. Akad. Nauk, Moscow, 2001).
12. T. A. Lapushkina, S. V. Bobashev, R. V. Vasil'eva, *et al.*, Zh. Tekh. Fiz. **72** (4), 23 (2002) [Tech. Phys. **47**, 397 (2002)].

13. S. V. Bobashev, A. V. Erofeev, T. A. Lapushkina, *et al.*, in *Proceedings of 32nd AIAA Plasmadynamics and Lasers Conference, 4th Weakly Ionized Gases Workshop, Anaheim, 2001* (AIAA Pap. **2001-2878**, 2001).

14. L. M. Biberman, V. S. Vorob'ev, and I. T. Yakubov, *Kinetics of Nonequilibrium Low-Temperature Plasmas* (Nauka, Moscow, 1982; Consultants Bureau, New York, 1987).

15. R. V. Vasil'eva, A. L. Genkin, V. L. Goryachev, A. V. Erofeev, A. D. Zuev, D. N. Mironov, A. S. Remennyĭ, and N. A. Silin, *Low-Temperature Plasma of Inert Gases with Nonequilibrium Ionization and MHD Generators* (Nauka, St. Petersburg, 1991).

16. N. N. Ogurtsova, I. V. Podmoshenskiĭ, and N. I. Demidov, Opt. Mekh. Prom. **1** (1) (1960).

17. L. A. Vulis, A. L. Genkin, and V. A. Fomenko, *The Theory and Analysis of Magnetohydrodynamic Flows in Channels* (Atomizdat, Moscow, 1971).

18. Yu. P. Raĭzer, *The Physics of Gas Discharge* (Nauka, Moscow, 1987).

Translated by N. Mende

Technical Physics, Vol. 48, No. 2, 2003, pp. 185–189. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 51–55.

GASES AND LIQUIDS

Cluster Model of a Gas

L. I. Kurlapov

Al-Farabi State University of Kazakhstan,
ul. Vinogradova 95, Almaty, 480012 Kazakhstan

Received June 4, 2002; in final form, August 19, 2002

Abstract—It is assumed that gas properties are defined by the motion and interaction not only of individual molecules but also of their clusters, which consist of several molecules. From the exponential cluster size distribution, the basic relationships for the parameters of cluster gas mixtures are derived, a computational scheme is suggested, and the cluster concentration and the compressibility factor for a number of gases are calculated. Using the example of the viscosity coefficient of oxygen, it is demonstrated that the cluster model fits the experimental pressure dependence of the viscosity if the variation of the cluster composition and partial contributions of cluster components to the viscosity are taken into account. *© 2003 MAIK "Nauka/Interperiodica".*

The kinetic theory of gases explains the properties of gases by the motion and interaction of individual structure elements to which dynamics laws are applicable. The basic difficulty here is to find an object that is such a structure element. In gases, these are molecules, which are thought to be responsible for the chemical properties. When the motion of molecules is described in terms of dynamics laws, they are viewed as compact and stable objects. However, molecules are complex thermodynamic systems and may aggregate into stable and unstable clusters or macromolecules (see, e.g., [1]). In the cluster model of gases, the behavior of real gases, which diverges from that of a perfect gas, is explained by the formation or dissociation of clusters according to the variation of macroparameters.

Physically this divergence is explained by the variation of the number of structure elements, which decreases upon clustering and grows upon cluster dissociation. In the equation of state, this is reflected by the variable number of moles and the variable molar mass of a cluster mixture, since a gas should be considered as a multicomponent mixture consisting of dimers, trimers, and higher order clusters. The presence of many components and the variability of the cluster composition significantly affect the nonequilibrium properties of a gas, which shows up, in particular, in the temperature and pressure dependences of the transport coefficients.

In this work, we derive basic relationships for the parameters of gas cluster mixtures, suggest a computational scheme, and calculate a number of equilibrium and transport parameters.

An equilibrium thermostatically controlled cluster gas can be described by a statistical ensemble with the canonical particle energy distribution [2, 3]. The potential energy of a cluster consisting of g molecules that is averaged over configurations is proportional to the number of molecules incorporated:

$$E_g = GgkT, \tag{1}$$

where G is a normalizing factor, k is the Boltzmann constant, and T is temperature.

Considering the amount n_g of g-dimensional clusters as the population of an appropriate energy level in the canonical ensemble, one can write

$$n_g = A\exp\left[-\frac{Eg}{kT}\right], \tag{2}$$

where A is a normalizing preexponential.

In view of (1), this distribution takes the form

$$n_g = A\exp(-Gg). \tag{3}$$

The normalizing factor A can be expressed through the amount of monomers (individual moving molecules, for which $g = 1$):

$$A = n_1\exp(G), \tag{4}$$

where n_1 is the number of monomers per unit volume.

Substituting (4) into (3) yields the exponential cluster size distribution:

$$n_g = n_1\exp[(-G(g-1)]. \tag{5}$$

Dividing this expression by the total amount of clusters $n^{(c)} = \sum_{g=1}^{r} n_g$ (r is the size of largest clusters taken into consideration) yields the cluster fraction (concentration) distribution over sizes:

$$C_g^{(c)} = C_1^{(c)}\exp[-G(g-1)], \qquad C_g^{(c)} = \frac{n_g}{n^{(c)}}, \tag{6}$$

where $C_1^{(c)}$ is the concentration of molecules (monomers).

1063-7842/03/4802-0185$24.00

In the cluster model, a gas is considered as a multicomponent mixture of clusters. The composition of the mixture can be conveniently expressed through macroparameters such as the average molar mass

$$\langle M\rangle = \sum_{g=1}^{r} C_g^{(c)} M_g, \tag{7}$$

where M_g is the molar mass of a g-dimensional cluster component and $\langle M\rangle$ is the average molar mass of the cluster mixture.

Bearing in mind that the molar mass is defined as the mass of structure components taken in an amount equal to the Avogadro number, one can write the molar mass of a g-dimensional cluster component as

$$M_g = N_A m_i g. \tag{8}$$

Since the mass of a g-dimensional cluster is the sum of the masses m_i of g individual molecules, $M_g = M_1 g$.

Structure elements responsible for chemical properties (of molecules) can be distinguished from those producing gas pressure by introducing the compressibility factor. With this factor, the real gas equation is given by

$$p = z n^{(n)} kT, \tag{9}$$

where $n^{(n)}$ is the total amount of both free and bound (clustered) molecules and z is the compressibility factor (see, e.g., [3; p. 202, 259]).

In the cluster model, each cluster component is viewed as an independent participant of processes. The gas equation for it can be written in the form

$$p_g = \frac{m_g}{V(1-b)M_g} RT, \tag{10}$$

where p_g is the partial pressure of a g-dimensional cluster component, m_g is the mass of the g-dimensional cluster, and b is the correction for the relative (specific) volume of the particles.

In terms of the partial mass density, this equation appears as

$$p_g = \frac{\rho_g}{(1-b)M_g} RT, \tag{11}$$

where ρ_g is the partial mass density of a g-dimensional cluster.

The mass density of a cluster component can be expressed through the mass m_i of one molecule:

$$\rho_g = n_g m_i g. \tag{12}$$

Thus, the partial pressure is given by

$$p_g = \frac{n_g}{(1-b)} kT. \tag{13}$$

For a cluster mixture, the gas pressure is the sum of the partial pressures of cluster components:

$$p = \frac{1}{(1-b)} kT \sum_{g=1}^{r} n_g. \tag{14}$$

The total pressure of a gas is convenient to express through the ratio of the number of clusters to the total number of molecules:

$$p = \frac{n^{(n)}}{(1-b)} kT \sum_{g=1}^{r} C_g^{(n)}, \tag{15}$$

$$C_g^{(n)} = \frac{n_g}{n^{(n)}} = \frac{n_g}{\sum_{g=1}^{r} g n_g}, \tag{16}$$

where r is the size of largest clusters taken into consideration.

From (9) and (15), we find a formula for the compressibility factor:

$$z = \frac{1}{(1-b)} \sum_{g=1}^{r} C_g^{(n)}. \tag{17}$$

It is seen that the compressibility factor can be calculated if the concentrations $C_g^{(n)}$ of clusters of various size are known. Computational schemes usually use concentrations $C_g^{(c)}$, from which $C_g^{(n)}$ can be found:

$$\begin{aligned} C_g^{(n)} &= \frac{n_g}{n^{(n)}} = \frac{n_g}{\sum_{g=1}^{r} g n_g} \\ &= C_g^{(c)} \frac{n^{(c)}}{\sum_{g=1}^{r} g n_g} = C_g^{(c)} = \frac{1}{\sum_{g=1}^{r} g C_g^{(c)}}, \end{aligned} \tag{18}$$

where $n^{(c)}$ is the total amount of clusters.

In this work, the compressibility factor is calculated by the formula

$$z = \frac{1}{(1-b)} \sum_{g=1}^{r} \frac{C_g^{(c)}}{\sum_{g=1}^{r} g C_g^{(c)}}, \tag{19}$$

where b is the relative correction for the partial volume of the particles. In terms of the effective molecule-collision diameter σ, this correction is given by

$$b = \frac{2\psi}{3} n^{(n)} \pi \sigma^3, \tag{20}$$

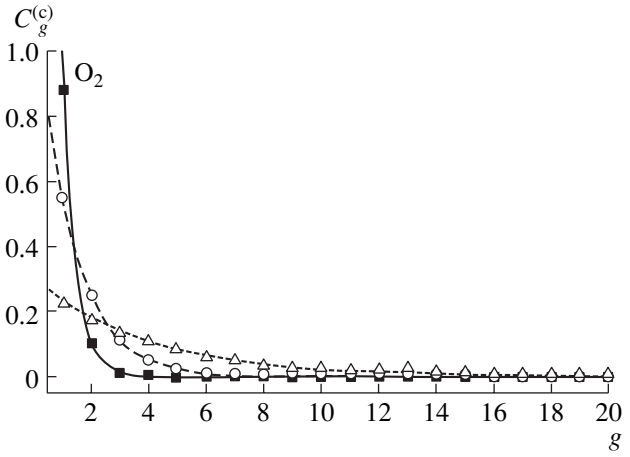


Fig. 1. Concentration $C_g^{(c)}$ of cluster components vs. cluster dimension g (expressed in numbers of molecules) for oxygen pressure p = (■) 6, (○) 40, and (△) 90 MPa at 300 K. The calculation was made for g = 1–20 with the effective collision diameter σ_{11} = 0.308 nm evaluated from the viscosity coefficient. Curves are plotted by the formula $C_g^{(c)} = C_1^{(c)} \exp[-G(g-1)]$.

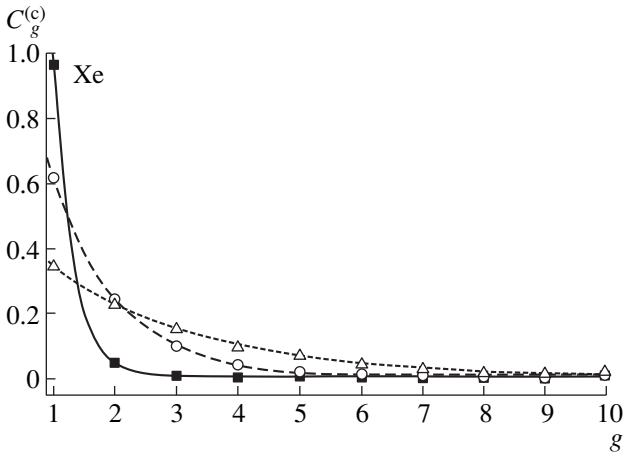


Fig. 2. Concentration $C_g^{(c)}$ of cluster components vs. cluster dimension g (expressed in numbers of molecules) for xenon pressure p = (■) 0.5, (○) 4, and (△) 6 MPa at 300 K. The calculation was made for g = 1–10 with the effective collision diameter σ_{11} = 0.411 nm evaluated from the viscosity coefficient. Curves are plotted by the formula $C_g^{(c)} = C_1^{(c)} \exp[-G(g-1)]$.

where ψ is the packing parameter taking into account the fact that the (specific) particle volume may change upon clustering.

The above relationships give a set of equations for the concentration $C_g^{(c)}$:

$$C_1^{(c)} \sum_{g=1}^{r} \{g \exp[-G(g-1)]\} - \frac{\rho RT}{pM_1(1-b)} = 0,$$

$$C_1^{(c)} \left(1 + \sum_{g=2}^{r} \exp[-G(g-1)]\right) - 1 = 0, \qquad (21)$$

$$C_1^{(c)} \exp[-G(g-1)] - C_g^{(c)} = 0, \quad g = 1-r.$$

The results of calculation for specific cases performed with this scheme are presented in Figs. 1 and 2 and in Tables 1 and 2. In the calculations, we used published data for the density or specific volume. The effective diameters were evaluated from reference data for viscosity coefficients (for oxygen and xenon [4, 5]) or self-diffusion coefficients [4] with allowance for their temperature dependences by using the kinetic theory formulas [6–8]. As follows from the figures, the size of clusters may be rather large at high pressures. Their presence influences the compressibility factor. Specifically, if the fraction of xenon molecules combined into clusters is high, the compressibility factor drops to less than unity. Table 2 shows that the compressibility factors calculated are in good agreement with the reference data.

In the cluster model, a chemically pure gas is considered as consisting of many cluster components. The reason for the temperature or pressure dependence of

Table 1. Cluster concentrations $C_s^{(c)}$ for carbon dioxide at a pressure 2 MPa and various temperatures

Cluster dimension, g	Temperature, K			
	300	400	600	700
1	0.8172	0.9144	0.9696	0.9796
2	0.494	0.0783	0.0295	0.0200
3	0.0273	0.0067	8.9918×10^{-4}	4.0907×10^{-4}
4	0.0050	5.744×10^{-4}	0	0
5	9.1189×10^{-4}	0	0	0
6	1.6666×10^{-4}	0	0	0

Note: Calculations were made for g = 1–6 with effective collision diameters evaluated from the self-diffusion coefficients [4].

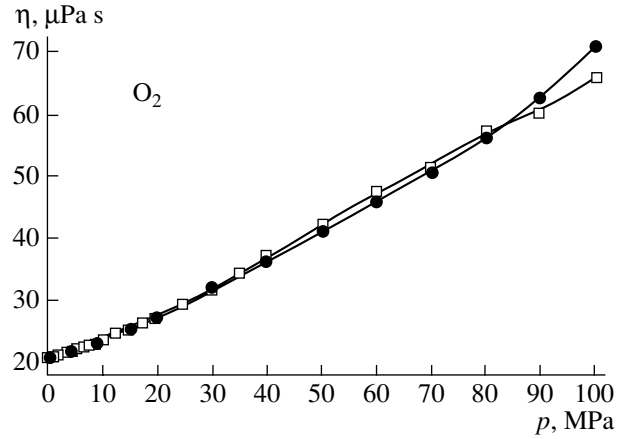


Fig. 3. Pressure dependence of the viscosity coefficient η at 300 K: (squares) smoothed tabulated data [4] and (circles) calculation by formula (22) including clusters with the viscous effective collision diameter σ_{11} = 0.308 nm. The cluster dimension g = 1–20. The dashed line shows the constant value η = 20.7 μPa s for a Boltzmann gas [4].

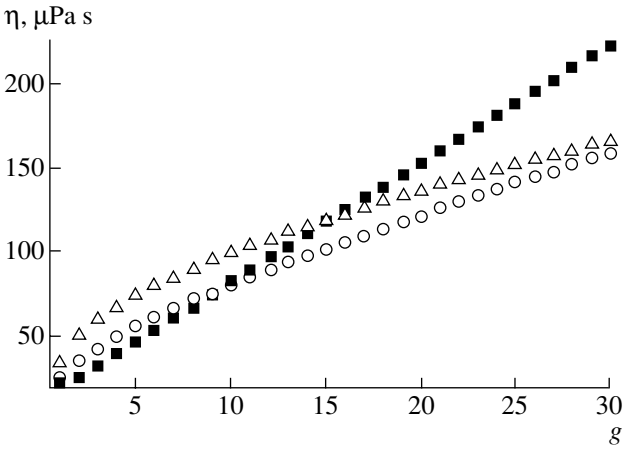


Fig. 4. Partial viscosity coefficient of a cluster component vs. cluster dimension. The values of p are the same as in Fig. 1. T = 300 K.

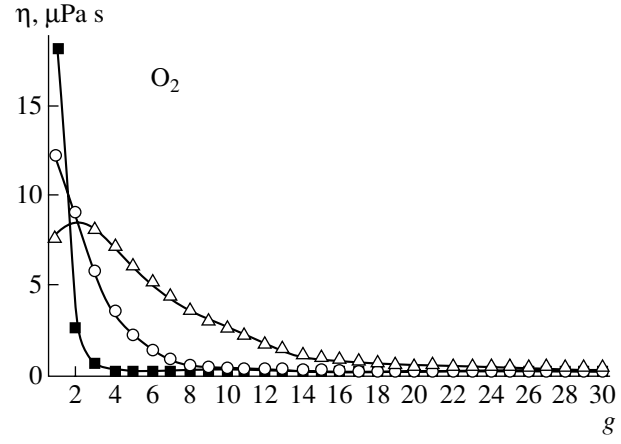


Fig. 5. Fraction of clusters vs. their dimension. The values of p are the same as in Fig. 1. T = 300 K.

the transfer coefficients is that a temperature decrease or a pressure increase favors the formation of larger clusters, which make a greater contribution to transfer processes. Specifically, the pressure dependence of the viscosity coefficient is associated with the formation of large clusters with a high molar mass. The formulas of the kinetic theory [6, 7] for the viscosity of a multicomponent mixture can also be applied for a cluster mixture. With Enskog corrections taken into account, the formula for the viscosity coefficient of the mixture has the form

$$\eta = \sum_{\alpha=1}^{s} \frac{x_\alpha h\sqrt{T}}{\sum_{\beta=1}^{s} x_\beta \sigma_{\alpha\beta}^2 \sqrt{\frac{2M_{\beta\alpha}}{M_\alpha}} (5M_{\alpha\beta}\Omega_{\alpha\beta}^{1,1*} + 3M_{\beta\alpha}\Omega_{\alpha\beta}^{2,2*})\chi_{\alpha\beta}}, \quad (22)$$

where $h = 8009 \times 10^{-29}$ J$^{1/2}$ K$^{-1/2}$ kmol$^{1/2}$ is the dimensional coefficient.

With allowance for collisional momentum transfer within the effective diameter, the Enskog correction $\chi_{\alpha\beta}$ has the form [9]

$$\chi_{\alpha\beta} = \frac{1}{S^3(1+s_{\alpha\beta})} \times \left\{ S^2 + 3\frac{\sigma_{\alpha\alpha}\sigma_{\beta\beta}}{\sigma_{\alpha\alpha}+\sigma_{\beta\beta}} S S_2 + 2\left(\frac{\sigma_{\alpha\alpha}\sigma_{\beta\beta}}{\sigma_{\alpha\alpha}+\sigma_{\beta\beta}}\right)^2 S_2^2 \right\},$$

$$S_2 = \frac{\pi n}{6} \sum_{\beta=1}^{s} \chi_\beta \sigma_{\beta\beta}^2, \quad (23)$$

$$S = 1 - \frac{\pi n}{6} \sum_{\beta=1}^{s} \chi_\beta \sigma_{\beta\beta}^2,$$

$$s_{\alpha\beta} = \frac{\sigma_{\alpha\beta}}{\tau_\alpha}\sqrt{\frac{m_\alpha}{3kT}}.$$

In cluster mixtures consisting of molecules of one sort, the cluster dimension g plays the role of the component number and summation is over the sizes of clusters taken into account.

Figures 3–5 show the results of calculation by formula (22). The effective molecule diameter, which also enters into the formula for the equilibrium parameters (compressibility factor and cluster concentration), is found by the same formula. Here, reference data for the viscosity coefficient of a specific gas are taken under conditions when the clustering effect can be neglected (rarefied, or Boltzmann, gas), which is usually observed under atmospheric pressure.

The effect of clustering on the transport parameters is demonstrated with Fig. 3, where the pressure dependence of the viscosity coefficient of oxygen is shown. The calculation is seen to be in good agreement with

Table 2. Compressibility factor at $T = 500$ K

Pressure p, MPa	Compressibility factor z			
	oxygen		xenon	
	reference data [5]	calculation by (19)	reference data [5]	calculation by (19)
0.1	1.0002	0.9997	0.9991	0.9994
1	1.0018	1.0023	0.9911	0.9909
2	1.0037	1.0036	1.0823	0.9822
3	1.0057	1.0056	0.9735	0.9735
4	1.0078	1.0078	0.9649	0.9649
5	1.0099	1.0099	0.9563	0.9563
6	1.0122	1.0122	0.9479	0.9479
8	1.0169	1.0169	0.9316	0.9315
10	1.0221	1.0221	0.9160	0.9159
12	1.0275	1.0275	0.9013	0.9012
16	1.0394	1.0395	0.8754	0.8754
20	1.0525	1.0526	–	
25	1.0706	1.0706		
30	1.0903	1.0903		
35	1.1115	1.1115		
40	1.1339	1.1339		
45	1.1575	1.158		
50	1.182	1.182		

the published data. Physically, this dependence is explained as follows. As is seen from formula (22), the total viscosity coefficient is the sum of the partial contributions from cluster components. This formula describes well the concentration dependence of the viscosity of a rarefied gas mixture consisting of chemically different components [7]. Under elevated pressures, a gas consisting of chemically identical molecules contains clusters in the form of dimers, trimers, etc. Within the cluster model, they are treated in terms of the formulas from the kinetic theory (specifically, in terms of formula (22)) as if they are independent mixture components. When colliding with molecules or lighter clusters, heavy clusters have a longer free path (the effect of velocity persistence after collision); therefore, their contribution to the momentum flux and the mixture viscosity grows. Moreover, the relationships between the masses and mean thermal velocities are such that heavy clusters transfer the momentum more effectively. This statement is supported by Fig. 4, which shows the partial viscosity coefficient vs. cluster dimension. The weaker dependence under high pressures is due to the fact that, when the concentration of heavy clusters is significant, they collide with each other more frequently and the effect of velocity persistence after collision is reduced. Thus, the viscosity of a cluster mixture increases with the fraction of heavy clusters. However, the fraction of heavy clusters drops exponentially with the cluster dimension (Fig. 5); therefore, their contribution to the viscosity is not so pronounced as follows from Fig. 4. The competition of these two mechanisms is responsible for the pressure dependence of the gas viscosity (Fig. 3).

REFERENCES

1. S. Ya. Khmel' and R. G. Sharafutdinov, Zh. Tekh. Fiz. **68** (8), 120 (1998) [Tech. Phys. **43**, 986 (1998)].
2. R. P. Feynman, *Statistical Mechanics: A Set of Lectures* (Benjamin, Reading, Mass., 1972; Mir, Moscow, 1975).
3. J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954; Inostrannaya Literatura, Moscow, 1961).
4. N. B. Vargaftik, *Tables of the Thermophysical Properties of Liquids and Gases*, 2nd ed. (Nauka, Moscow, 1972; Halsted, New York, 1975).
5. *Thermophysical Properties of the Technically Important Gases under High Temperatures and Pressures. Handbook*, Ed. by V. N. Zubarev, A. D. Kozlov, V. M. Kuznetsov, *et al.* (Énergoatomizdat, Moscow, 1989).
6. L. I. Kurlapov, Zh. Tekh. Fiz. **48**, 1302 (1978) [Sov. Phys. Tech. Phys. **23**, 731 (1978)].
7. L. I. Kurlapov, *Kinetic Theory of Irreversible Processes in Gases: Monograph* (Almaty, 2000).
8. L. I. Kurlapov, Zh. Tekh. Fiz. **46**, 1777 (1976) [Sov. Phys. Tech. Phys. **21**, 1034 (1976)].
9. J. M. Kincaid, M. Lopez de Haro, and E. G. D. Cohen, J. Chem. Phys. **79**, 4509 (1983).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 190–198. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 56–64.

GAS DISCHARGES, PLASMA

Simulations of Plasma Formation in the Cathode Sheath of an Efficient Excimer Lamp

A. N. Tkachev and S. I. Yakovlenko

Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia

Received April 22, 2002; in final form, August 15, 2002

Abstract—Based on the available experimental data and computer simulations, analytical approximations of the quantities characterizing electron multiplication in the cathode sheath are proposed. The critical electric field is found above which runaway electrons are observed. Using the approximations proposed, the dependences of the plasma parameters (the electron and ion densities and currents and the electric field strength) on the distance from the cathode are analyzed. Simple formulas for the total current, the cathode sheath thickness, and the cathode potential drop as functions of the electric field on the cathode surface are derived. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In recent years, excimer lamps—the sources of spontaneous emission on the transitions of exciplex and excimer molecules—have attracted considerable interest [1–5]. In particular, excimer lamps on xenon dimers have been intensely developed [3–5]. To pump excimer lamps, barrier discharges, which allow one to obtain 172-nm radiation with an efficiency of higher than 50% [5], are most frequently used. Earlier, such a high efficiency was obtained by pumping high-pressure xenon by either an electron beam or an electron-beam sustained discharge [6–8].

The highest efficiency was attained with lamps in which a single-barrier discharge was ignited between a thin wire cathode located on the tube axis and an anode placed on the outer surface of a quartz tube. The xenon pressure was fairly high (~100–200 torr). Simulations [9] show that this pressure is indeed optimum for electron beam pumping. However, the kinetics of a plasma pumped by an electron beam differs significantly from the kinetics of a discharge plasma (see, e.g., [10]). In this context, it is of interest to study the discharge physics in dense xenon. The discharge characteristics are, to a great extent, governed by the processes occurring in the cathode sheath. This paper is devoted to studying the plasma parameters in the cathode sheath.

The following problems are considered. First, we simulate electron multiplication near the cathode, where the screening of the electric field by the plasma is of minor importance. Then, based on the simulation results and available experimental data, we obtain the Townsend ionization coefficient, drift velocity, electron mean energy, and other ionization characteristics as functions of the electric field. These data are used to solve the set of equations describing the current generation in the cathode sheath. Based on the simulation results, simple analytical expressions for the cathode sheath parameters are derived.

1. SIMULATIONS OF IONIZATION

1.1. Simulations of Electron Multiplication

The problem is solved in cylindrical geometry. The grounded anode is a cylinder of radius $R_{max} \sim 1$ cm. The cathode, which is a wire of radius $r_0 = 0.5$ mm located on the axis of the anode cylinder, is at the potential $U_0 \sim -(0.5–3.0)$ kV.

In the absence of plasma, the electric field depends on the distance r from the wire axis as

$$E(r) = E_0\frac{r_0}{r}, \quad E_0 \equiv E(r_0) = \frac{-U_0}{\ln(r_0/R_{max})}\frac{1}{r_0}. \quad (1)$$

For the above parameters and in the absence of plasma, the electric field near the cathode is in the range $E_0 \sim 3–20$ kV/cm.

Electron multiplication in electric field (1) was simulated by the particle method. The electron motion was computed over a certain radial interval $r_1 < r < r_2$ ($r_1 \geq r_0$, $r_2 \leq R_{max}$), which was divided into ten to one hundred cells. The numerical grid obtained was used to acquire the statistics for the parameters under study (the ionization rate, electron velocity, electron mean energy, etc.). The electrons were injected at a certain rate from the point $r = r_1$ into the computation region. The energies of the injected electrons ε were distributed by Poisson's law, $P_1 = \varepsilon_0\exp(-\varepsilon/\varepsilon_0)$, with the characteristic energy ε_0, which was set at $\varepsilon_0 = 0.2$ eV. During the time intervals between the electron injection events, the evolution of all the electrons in the computation region was traced and statistics was gained. For this purpose, the time step Δt was chosen such that, for any electron energy below

1063-7842/03/4802-0190$24.00

3 keV, the electron displacement during this time step did not exceed 30% of the mean free path (commonly, under the above conditions, the time step was $\Delta t = 0.5$ fs). At each time step, the equations of motion for all the electrons (both newly injected and already residing in the computation region) in the electric field were solved and the electrons were set at new positions.

The electrons that reached the boundaries of the computation region ($r \geq R_{max}$ or $r \leq R_{min}$) were excluded from consideration (disappeared). The remaining electrons underwent collisions with the collision probability determined by Poisson's law

$$P_2(\Delta l) = 1 - \exp(-\Delta l/L),$$

where Δl is the distance passed by an electron for the last time step and L is the electron mean free path.

To determine whether or not the collision occurs, a random number u distributed uniformly over the interval [0, 1] was generated. The collision was considered to occur if, for a given Δl, we obtained $u < P_2$. To generate random numbers, we used a multiplicative generator.

The mean free path L is expressed via cross sections as

$$L^{-1} = N_{Xe}[\sigma_{el}(\varepsilon) + \sigma_{ex}(\varepsilon) + \sigma_i(\varepsilon)],$$

where ε is the electron energy and σ_{el}, σ_{ex}, and σ_i are the cross sections for elastic scattering, excitation to the first resonance level, and ionization, respectively (see Section 1.2).

Then, by using the ratio among the cross sections and with the help of the random number generator, we determined which one of the collision events had occurred: elastic scattering, excitation, or ionization.

In the case of elastic scattering, the electron velocity was rotated by angles φ and χ with respect to the previous electron velocity. For the angular distribution, we used an expression proposed in [11] that models the collision of an electron with a screened point charge at low and moderate energies ε,

$$f(\varphi, \chi)d\varphi d(\cos\chi) = \frac{\tilde{\varepsilon}}{4\pi\ln(1+\tilde{\varepsilon})}\frac{d\Omega}{\left(\tilde{\varepsilon}\sin^2\frac{\chi}{2} + 1\right)}, \quad (2)$$

$$\tilde{\varepsilon} = \frac{\varepsilon}{\varepsilon_1},$$

where ε_1 is the characteristic energy. In subsequent calculations, this energy is set at 1 eV.

To determine the scattering angles distributed in accordance with this expression, pairs of random numbers (u_1 and u_2) distributed uniformly over the interval [0, 1] were generated. The scattering angles were calculated using the expressions

$$\cos\chi = 1 - 2\frac{(1+\tilde{\varepsilon})^{u_1} - 1}{\tilde{\varepsilon}}, \quad \varphi = 2\pi u_2.$$

In the case of excitation, the electron energy was decreased by the excitation energy, the velocity direction was changed in the same way as with elastic scattering, and the number of the excited atoms in the given spatial cell was increased by one.

In the case of ionization, a new electron appeared and the number of ions in the given spatial cell was increased by one. The spatial coordinates of a newly born electron were assumed to be the same as those of the parent atom. The energy ε of the newly born electron was assumed to be distributed according to the Thomson model: $F(\varepsilon) = I/(\varepsilon - I)^2$. The rest of the energy was given to the primary electron. The velocities of both the primary and newly born electrons were distributed by formula (2), as in the case with elastic collision.

In the course of computations, statistics was gained to determine the ionization frequency in different spatial cells, the mean electron energy and the mean radial velocity as functions of radius, the electron energy distribution function (EEDF), etc.

The computation time was chosen long enough not only for the electron multiplication process to be established, but also a reliable statistics for the computed quantities to be gained.

1.2. Approximation of the Electron-Impact Cross Sections

The elastic scattering cross section was calculated by interpolating the table data. The data for energies of 0.01–20 eV were taken from [12]; for energies of 20–36 eV, they were taken from [13, Fig. 3]; and for energies in the range from 36 eV to 3 keV, it was assumed that the cross section decreased with energy as $1/\varepsilon$. The ionization cross section in the range 0–3 keV was calculated by interpolating the table data from [14].

For the sake of convenience, we derived approximation formulas that generalized the available information about the dependence of various cross sections on the energy of the incident electron.

The elastic scattering cross section in the energy range from 0.01 eV to 3 keV was approximated with an accuracy of 20% by the expression

$$\sigma_{el}(\varepsilon) = \left(\frac{2\varepsilon^2}{1 + 0.0045\varepsilon^3} + \frac{12\exp(-6.6\varepsilon^{1.1})}{\sqrt{\varepsilon}}\right) \times 10^{-16} \text{ cm}^2,$$

where ε is the energy of the incident electron in electronvolts.

In the same energy range, the single-ionization cross section can be approximated with an accuracy of 15%

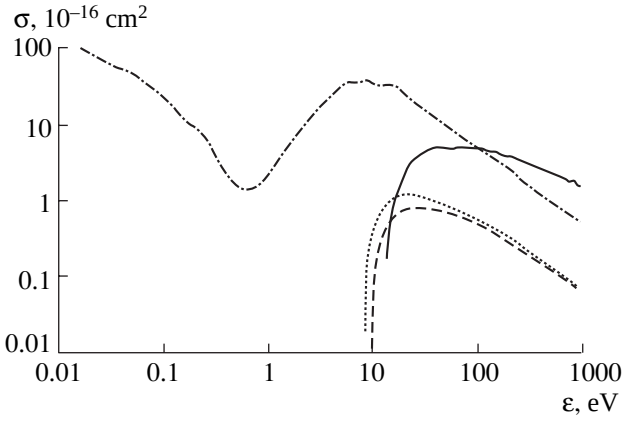


Fig. 1. Total cross section for electron collisions with Xe atoms vs. incident electron energy. The solid curve is for ionization; the dotted and dashed curves are for the excitation to the first and second resonance states, respectively; and the dashed-and-dotted curve is for elastic collisions.

by the expression

$$\sigma_i(\varepsilon) = \frac{16(\varepsilon - I)^{1.25}}{1 + 0.0032(\varepsilon - I)^{1.77}} \times 10^{-18}\ \text{cm}^2, \quad \varepsilon \geq I,$$

where $I = 12.13$ eV is the ionization energy of a xenon atom and the electron energy ε is in electronvolts.

For the excitation cross section in the energy range 0–3 keV, we used the following simplified approximation:

$$\sigma_{ex}(\varepsilon) = 4.24 \times 10^{-16}\ \text{cm}^2 \frac{(23 - E_1)(\varepsilon - E_1)}{(\varepsilon + 23 - 2E_1)^2} \quad (3)$$

$(\varepsilon > E_1$, energies are in electronvolts$)$.

Formula (3) approximates the known Drawin formula (see [15, Fig. 9.4]) for the transition from the ground state to the first resonant state $5p^5 6s$ ($J = 1$),

$$\sigma(\varepsilon) = 4\pi a_0^2 \left(\frac{Ry}{E_1}\right)^2 f \times \left\{ \frac{u}{(u+1)^2} \ln[1.25(u+1)] + \frac{0.3}{u+1} \right\},$$

$$u = \frac{\varepsilon - E_1}{E_1},$$

where the energy $E_1 = 8.437$ eV and the transition oscillator strength is $f = 0.26$. The parameter values in formula (3) were chosen such that the integral of the cross section and its maximum position match the integral and the maximum position in the case of the Drawin approximation.

In some runs, to estimate the effect of the excitation processes on the local ionization parameters, we also took into account the excitation of Xe atoms to the second resonant state $5p^5 6s'[3/2]_1$ with the cross section

$$\sigma_{ex}^{(2)}(\varepsilon) = 2.9 \times 10^{-16}\ \text{cm}^2 \frac{(30 - E_2)(\varepsilon - E_2)}{(\varepsilon + 30 - 2E_2)^2}$$

$(\varepsilon > E_2$ energies are in electronvolts$)$.

This formula was derived in the same way as formula (3), by fitting to the Drawin approximation with the parameters $E_2 = 9.57$ eV and $f = 0.27$. The energy dependences of the cross sections used in the calculations are shown in Fig. 1.

Ionization rates. Based on the above approximations, we calculated the rates of the collision-induced transitions, $k = \int \sigma(\varepsilon)(2\varepsilon/m_e)^{1/2} f(\varepsilon) d\varepsilon$, where $f(\varepsilon)$ is the EEDF.

Figure 2a presents the rates of the collision-induced transitions calculated for the case of a Maxwellian EEDF

$$f(\varepsilon) \equiv f_M(\varepsilon) \equiv \frac{2\sqrt{\varepsilon}}{\sqrt{\pi} T_e^{3/2}} \exp(-\varepsilon/T_e),$$

where m_e is the electron mass and T_e is the electron temperature. It can be seen that significant excitation and ionization occur at a mean electron energy of higher than 1 eV.

However, a mean electron energy of several electronvolts is actually insufficient for efficient ionization. This fact contradicts the results of simulations of electron multiplication by the particle method. Calculations show (see below) that the mean electron energy exceeds 1 eV already at $E \sim 100$ V/cm. However, in reality, ionization begins to dominate over excitation at significantly higher electric fields. The point is that, near the cathode surface, the electron energy distribution is closer to a Druyvesteyn distribution than to a Maxwellian one.

The result of averaging the cross sections using the Druyvesteyn distribution

$$f(\varepsilon) = f_D(\varepsilon) \equiv \frac{2}{\Gamma(3/4)} \left(\frac{3m_e}{m_{Xe}}\right)^{3/4} \sqrt{\frac{\varepsilon}{\varepsilon_0}} \exp\left(-\frac{3m_e \varepsilon^2}{m_{Xe} \varepsilon_0^2}\right)$$

is illustrated in Fig. 2b. Here, $\Gamma(3/4) = 1.225$; m_{Xe} is the Xe atomic mass; and $\varepsilon_0 = eE/(\sigma_{el} N_{Xe})$ is the energy acquired by an electron in the electric field E over the distance $l = 1/(\sigma_{el} N_{Xe})$, which the electron travels between two successive elastic collisions.

At a given density of Xe atoms, the frequencies of ionization, $\nu_i = k_i N_{Xe}$, and excitation, $\nu_{ex} = k_{ex} N_{Xe}$, averaged using a Druyvesteyn distribution, appear to be functions of the electric field E only. It is seen that the

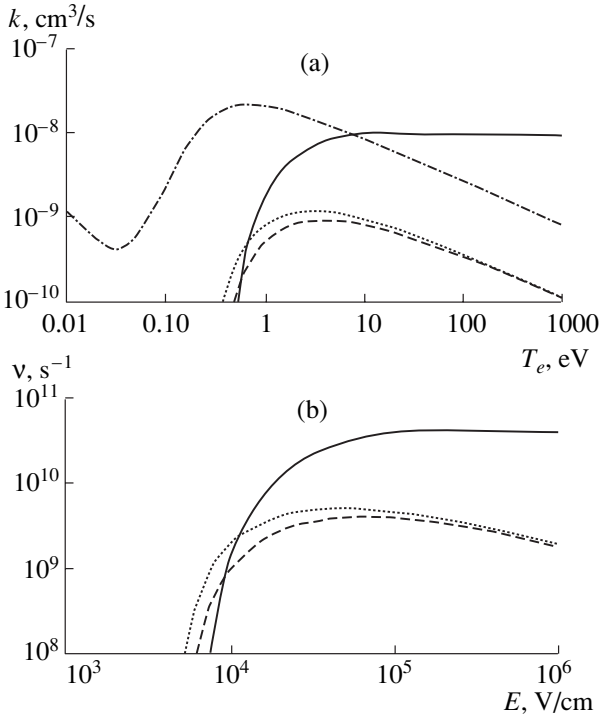


Fig. 2. (a) Rate constants of the electron-impact ionization and excitation of Xe atoms vs. electron temperature and (b) the ionization and excitation frequencies vs. electric field. The solid curves show the rate constant k_i and the ionization frequency ν_i (an ionization threshold of 12.13 eV), the dotted curves show the rate constant k_{ex1} and the frequency ν_{ex1} of the excitation to the $5p^5 6s[3/2]_1$ state (an excitation threshold of 8.437 eV), the dashed curves show the rate constant k_{ex2} and the frequency ν_{ex2} of the excitation to the $5p^5 6s'[3/2]_1$ state (an excitation threshold of 9.57 eV), and the dashed-and-dotted curve shows the elastic collision rate constant k_{el}.

high ionization efficiency, $\nu_{ex} < \nu_i \sim 10^{10}$ s^{-1}, is achieved at $E \sim 10^4$ V/cm.

1.3. Calculated Results and Approximations

Electron multiplication. In calculations, the xenon density was assumed to be relatively high, $N_{Xe} = 4.4 \times 10^{18}$ cm^{-3}. Intense electron multiplication and electric field screening take place in a narrow layer, whose thickness is much smaller than the cathode diameter. Generally speaking, the dependence of the local ionization parameters on the electric field should be determined by the modeling of the multiplication process in plane geometry (provided that the external electric field is uniform). In calculations of the ionization parameters, plane geometry was provided by choosing a sufficiently large cathode radius of curvature, $r_0 = 10^3$ mm, which is larger than all the other characteristic dimensions. The typical calculated results are shown in Fig. 3.

Simulations show that, the number of Xe atoms excited and ionized by electron impact exponentially

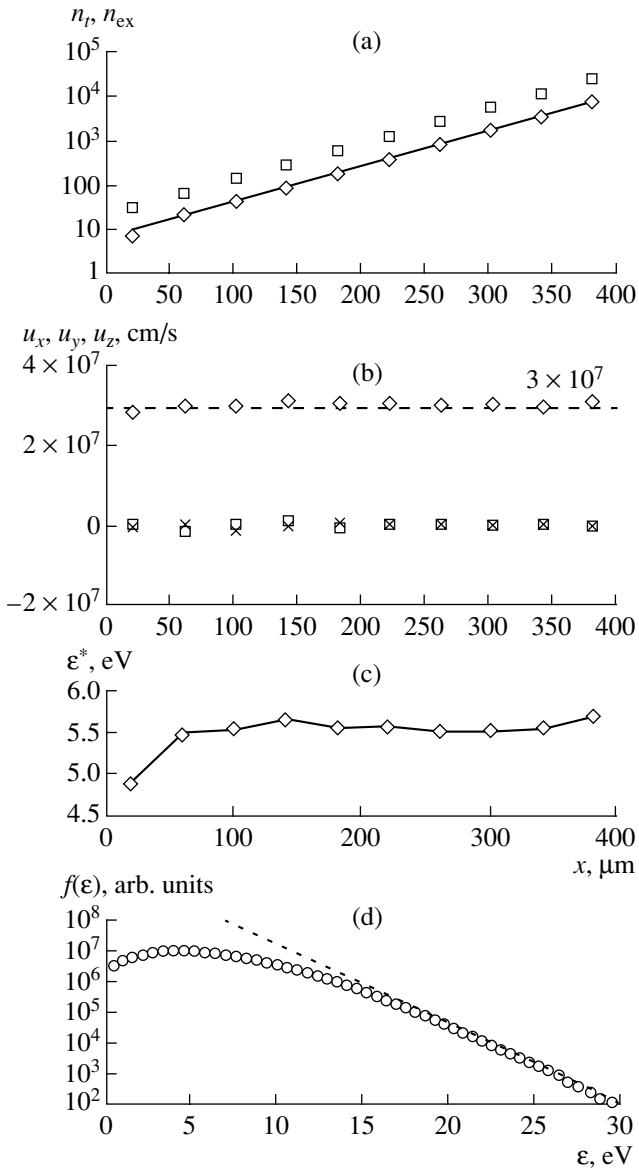


Fig. 3. Electron multiplication parameters calculated in plane geometry for an electric field of $E_0 = 9400$ V/cm and a xenon atom density of $N_{Xe} = 4.4 \times 10^{18}$ cm^{-3} (which corresponds to a pressure of 137 torr at room temperature): (a) the numbers of the created ions n_i (diamonds) and the excited atoms n_{ex} (squares) vs. distance x from the cathode; the solid curve is $n_i(x) = 7\exp(180x/\text{cm})$; (b) the mean electron velocity u_x (diamonds) along the x axis (along the electric field) and the mean electron velocities u_y (crosses) and u_z (squares) along the y and z axes, respectively, as functions of x; (c) the mean electron energy vs. x; and (d) the EEDF obtained by averaging over all the electrons in the interval x = 250–400 μm (the symbols show the simulation results and the dotted curve shows the dependence $f(\varepsilon) = 0.5\exp(-0.59\varepsilon)$.

increases with distance x from the cathode (Fig. 3a). The x-component of the mean electron velocity, u_x, directed along the electric field; the mean electron energy, ε^*; and the ionization frequency, ν_i, depend

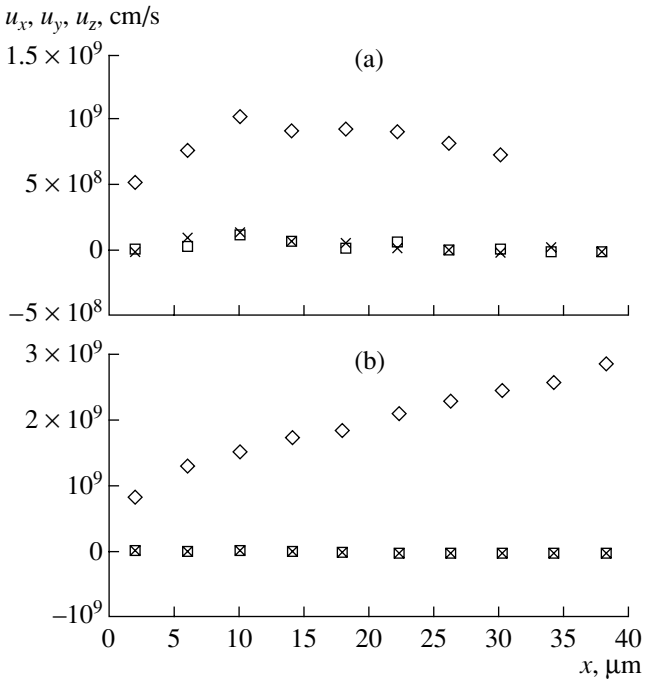


Fig. 4. Illustration of the transition to the electron runaway regime. The mean electron velocity vs. distance x from the cathode for an electric field of E_0: (a) 1.07×10^6 and (b) 2.13×10^6 V/cm: u_x (diamonds) is the mean electron velocity along the x axis (along the electric field) and u_y (crosses) and u_z (squares) are the mean velocities along the y and z axes, respectively.

slightly on the x coordinate (Figs. 3b–3d). The EEDF has a maximum at an energy lower than the ionization energy and decays exponentially at energies higher than the ionization energy (Fig. 3e).

This agrees with the Townsend theory, according to which the ionization frequency ν_i is related to the Townsend ionization coefficient α via the simple expression

$$\nu_i = \alpha u_e, \tag{4}$$

where u_e is the mean electron velocity along the electric field and the Townsend coefficient shows the number of the ionization events produced by an electron per unit path length. It is important that α and u_e mainly depend only on E/N and do not depend on the x coordinate. This fact is confirmed by the results of our simulations for the gas density that is one hundred times lower, namely, for $N_{Xe} = 4.4 \times 10^{16}$ cm^{-3}.

Note that the Townsend theory holds for electric fields that are lower than a certain critical value E_{cr}. At $E < E_{cr}$, the mean electron velocity continuously increases with distance; i.e., a runaway of electrons occurs. The transition to the runaway regime is illustrated in Fig. 4. At $E = 1.1 \times 10^6$ V/cm, the electron velocity is almost constant, and at $E = 2.1 \times 10^6$ V/cm, the mean electron velocity continuously increases.

According to the calculated results, the critical electric field for xenon is $E_{cr} \approx 10^4 (N_{Xe}/3.2 \times 10^{16}\ \text{cm}^{-3})$ V/cm. Below, we present approximations of the electron multiplication parameters corresponding to $E < E_{cr}$.

Approximations of the multiplication parameters. We performed a series of simulations (similar to those presented in Figs. 3 and 4) for different electric fields. Based on both these simulations and the experimental data, we proposed approximations for dependences of some quantities (that are used below in the set of equations describing the plasma formation in the cathode sheath) on E and N (Fig. 5).

The dependence of the Townsend coefficient on the electric field is in good agreement with the known approximation [13]

$$\alpha(E) = C_1 N \exp(-C_2 (N/E)^{1/2}), \tag{5}$$

where $C_1 = 2.03 \times 10^{-15}$ cm^2 and $C_2 = 2.014 \times 10^{-7}$ cm$^{3/2}$V$^{1/2}$ are the fitting parameters (Fig 5a).

The measured dependence of the drift velocity on the electric field (see [13] and Tables 3, 4 in [16]) was supplemented with the simulation results and approximated (Fig. 5b) by the expression

$$u_{d,e}(E) = \frac{4.18 \times 10^{22} E/N + 1.44 \times 10^{115} (E/N)^6}{1 + 4.06 \times 10^{106} (E/N)^{5.8}} + 4.18 \times 10^{21} \frac{E}{N} \exp\left(-6.82 \times 10^{-17} \frac{N}{E}\right), \text{ cm/s}. \tag{6}$$

Here and below, in the approximation formulas, N is in cm^{-3} and E is in V/cm.

The calculated ionization frequency ν_i (Fig. 5c) as a function of the electric field is approximated to within a factor of 3 by the product of expressions (5) and (6). The calculated mean electron energy and the experimental data [16] are described by the approximation (Fig. 5d)

$$\varepsilon^* = 0.026 + (5.5 - 0.026) \times \exp\left[-2.05 \times 10^{-9} \left(\frac{N}{E}\right)^{0.5}\right] + 4.4 \times 10^{14} \frac{E}{N}, \text{ eV}. \tag{7}$$

To solve the equations describing the plasma formation in the cathode sheath (see below), we also need to know the dependence of the ion drift velocity on the electric field. This dependence was obtained by approximating the experimental data of [15, 16]:

$$u_{d,i} = 3.3 \times 10^{18} \frac{1 + 2.1 \times 10^8 \left|\frac{E}{N}\right|^{12}}{1 + 1.056 \times 10^{15} \left|\frac{E}{N}\right|} \left|\frac{E}{N}\right|, \text{ cm/s}. \tag{8}$$

2. SIMULATIONS OF THE ELECTRON AND ION CURRENTS

2.1. Formulation of the Problem

Basic equations. In the cathode sheath, a plasma is formed that screens the electric field. The process of plasma formation is described by the momentum transfer equation, the continuity equations for electrons and ions, and Poisson's equation. In a steady state, in which the ionization and transfer processes in the cathode sheath comply with a given boundary electric field, the basic equations are as follows [11, 17, 18]:

$$\mathbf{u}_e = -eb_e\mathbf{E} - (D_e/N_e)\mathrm{grad}(N_e),$$
$$\mathrm{div}(N_e\mathbf{u}_e) = \nu_i N_e,$$
$$\mathrm{div}(\mathbf{E}) = 4\pi e(N_i - N_e);$$
$$\mathbf{u}_i = eb_i\mathbf{E} - (D_i/N_i)\mathrm{grad}(N_i),$$
$$\mathrm{div}(N_i\mathbf{u}_i) = \nu_i N_e.$$

Here, e is the electron charge, $\mathbf{u}$ is the mean velocity, b is the mobility, D is the diffusion coefficient, N is the density, and $\mathbf{E}$ is the electric field; the indices e and i stand for electrons and ions, respectively.

Let us turn to cylindrical geometry. In this case, we have grad $\longrightarrow d/dr$ and div $\longrightarrow (1/r)d(r\cdot)/dr$. Here, it is convenient to introduce the linear current densities

$$j_e = -2\pi e u_e N_e r, \quad j_i = -2\pi e u_i N_i r. \tag{9}$$

Then, using the charge conservation law in the form $j_e + j_i = j_0 = \mathrm{const}$ and expressing the particle densities via the current densities, we obtain the following set of equations

$$\frac{dj_e}{dr} = -2\pi e r \nu_i N_e,$$
$$\frac{dN_e}{dr} = \frac{j_e}{2\pi e r D_e} - \frac{b_e E N_e}{D_e},$$
$$\frac{dN_i}{dr} = \frac{j_e - j_0}{2\pi e r D_i} + \frac{b_i E N_e}{D_i}, \tag{10}$$
$$\frac{dE}{dr} = 4\pi e(N_i - N_e) - \frac{E}{r}, \quad j_i = j_0 - j_e.$$

The right-hand sides of Eqs. (10) are specified by the expressions

$$u_{d,i}(E) \equiv eb_i(E)E, \quad u_{d,e}(E) \equiv eb_e(E)E,$$
$$D_i(E) = T_i b_i(E), \quad D_e(E) = \varepsilon^*(E)b_i(E), \tag{11}$$

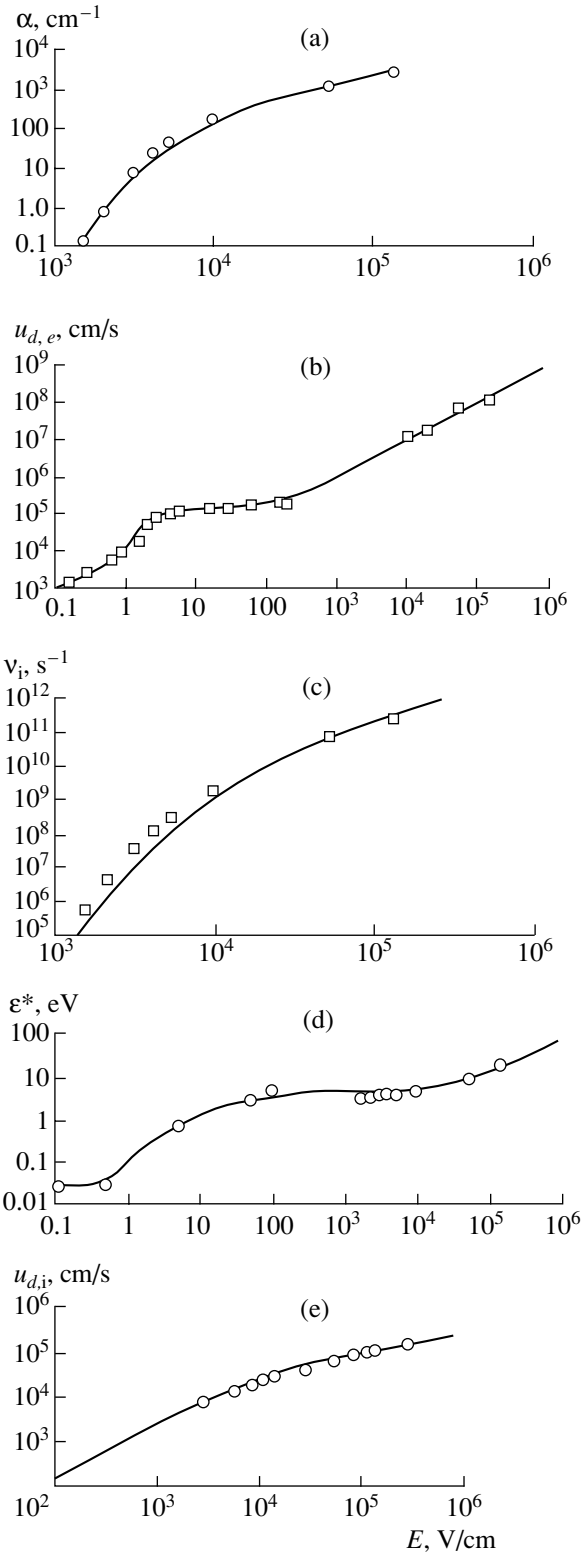


Fig. 5. Ionization and drift parameters vs. electric field at $N_{Xe} = 4.4 \times 10^{18}$ cm^{-3}: (a) Townsend coefficient approximated by formula (5) (solid curve) and obtained in simulations (symbols); (b) the electron drift velocity from [13, 15] (symbols, $E = (0.1–1.6) \times 10^4$ V/cm) and simulated (symbols, $E > 1000$ V/cm), the solid curve shows the approximation by formula (6); (c) the ionization frequency obtained in simulations (symbols) and from approximations (5) and (6) for the Townsend coefficient and the drift velocity, respectively (solid curve); (d) the characteristic electron energy obtained in simulations (symbols, $E = 10^3–10^5$ V/cm) and from approximation (7) (solid curve) and the data from [15] (symbols, $E = 0.1–100$ V/cm); and (e) the ion drift velocity approximated by formula (8) (solid curve) and the data from [15, 16] (symbols).

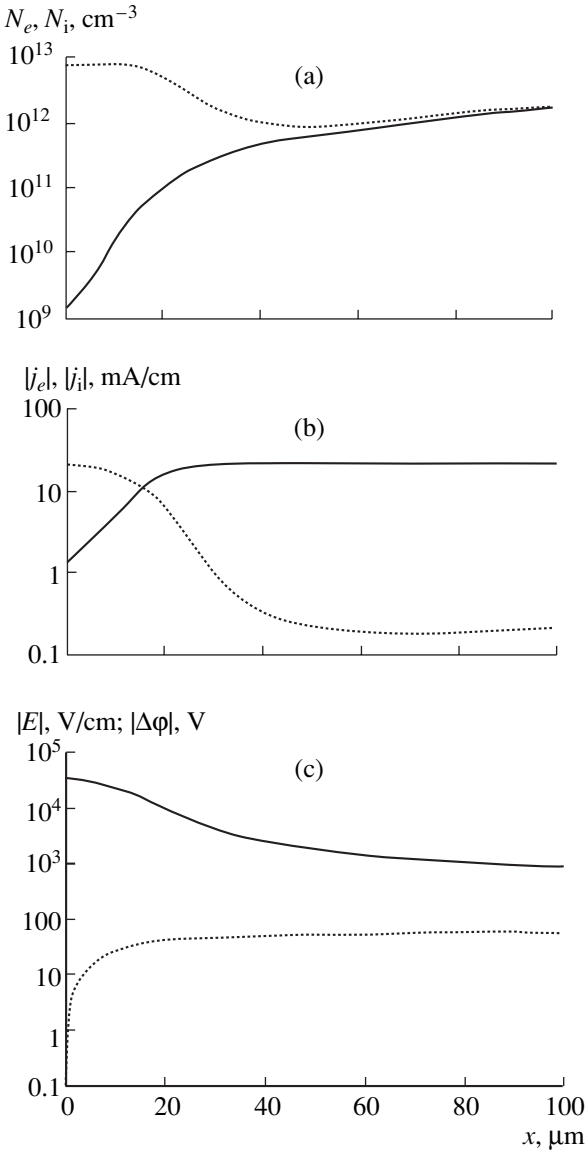


Fig. 6. Plasma parameters vs. distance from the cathode $x = r - r_0$ for $j_0 = 20$ mA/cm, $N_{Xe} = 4.4 \times 10^{18}$ cm^{-3}, $E_0 = 34$ kV/cm, $\xi = 0.07$, and $\zeta \approx 1.16$: (a) the electron (solid curve) and ion (dotted curve) densities, (b) the electron (solid curve) and ion (dotted curve) currents, and (c) the electric field (solid curve) and change in the electric field potential $\Delta\varphi(r) = \varphi(r_0) - \varphi(r)$ (dotted curve).

where $T_i = 0.03$ eV is the ion temperature and the quantities u_{di}, u_{de}, and ε^* are specified by formulas (5)–(7).

Boundary conditions. In order to solve the set of equations (10), boundary conditions should be imposed. In particular, at the large distances from the cathode, the ion and electron densities should be equal to each other:

$$N_e(r \longrightarrow \infty) = N_i(r \longrightarrow \infty). \quad (12)$$

In this study, we used the "shooting" method. The boundary conditions for the electron current, electron and ion densities, and electric field were imposed at the cathode surface, i.e., at $r = r_0$:

$$j_{e0} = j_e(r = r_0), \quad N_{e0} = N_e(r = r_0),$$
$$N_{i0} = N_i(r = r_0), \quad E_0 = E(r = r_0). \quad (13)$$

The value of j_{e0} was determined by using the preset values of the total current j_0 and the secondary electron emission coefficient ξ

$$j_{e0} = j_0\xi/(1 + \xi). \quad (14)$$

The boundary value of the ion density was set such that its gradient at the boundary was zero,

$$N_{i0} = j_0/(2\pi e r_0 b_i(E_0)E_0). \quad (15)$$

In the first approximation, the boundary value of the electron density was obtained from the condition that its gradient at the boundary comply with the Townsend multiplication,

$$N_{e0} = \zeta j_0[2\pi e r_0(b_e(E_0)E_0 + D_e(E_0)\alpha(E_0))], \quad (16)$$

where ξ is a fitting factor.

Then, the set of equations (10) was solved for given values of j_0 and E_0 with boundary conditions (13)–(16) at $\xi = 1$. Usually, the solution obtained did not satisfy condition (12). This condition was then satisfied by matching the ξ value. Note that, although the resultant ξ value insignificantly differed from unity, the solution was very sensitive to its exact value.

2.2. Simulation Results

Typical results from simulations of the plasma parameters in the cathode sheath are shown in Fig. 6. The total current $j_0 = 20$ mA/cm is taken from the experiment. Due to the screening of the electric field, the potential drop occurs in a very narrow sheath (see below); for this reason, the boundary electric field was set somewhat higher than follows from formula (1) for the electric field in the absence of plasma.

It can be seen from Fig. 6 that the physical processes can be described as follows. The intense electron multiplication occurs in a sheath of thickness $\Delta x \sim 30$ μm. Within the sheath, the electric field is screened by the ion charge. At $x \sim \Delta x$, the electric field decreases to a value at which the multiplication almost stops. Then, the electron and ion densities become nearly equal to each other and the current is mainly carried by electrons.

The potential drop across the sheath is relatively low, $\Delta\varphi_0 = |\varphi(r_0) - \varphi(r_0 + 100\ \mu m)| = 59$ V. It comprises a small fraction of the discharge voltage $U_0 \sim 0.5$–3 kV. The rest of the voltage, $U_0 - \Delta\varphi_0 \sim U_0$, falls on the quasineutral plasma, in which electron heating and, accordingly, thermal ionization play an important role. Here, we will not consider this region.

3. PLASMA PARAMETERS IN THE CATHODE SHEATH

3.1. Effect of the Cathode Electric Field

Assuming that the current is determined by the cathode sheath parameters, $j_0 = 2\pi e u_i N_{i0} r_0$, we express it as a function of the electric field at the cathode E_0. For this purpose, we determine the ion density from the condition for the electric field screening,

$$N_{i0}(E_0) = E_0/(4\pi e \Delta x(E_0)). \quad (17)$$

Let us estimate the sheath thickness via the Townsend coefficient,

$$\Delta x(E_0) = \ln(1/\xi)/\alpha(E_0). \quad (18)$$

As a result, we obtain

$$j_0(E_0) = \frac{u_{i,d}(E_0) E_0 r_0}{2\Delta x(E_0)}. \quad (19)$$

The quantities on the right-hand side of this expression are specified by formulas (4), (7), (17), and (18).

The potential drop across the cathode sheath can be estimated by the expression

$$\Delta\varphi_0(E_0) = E_0 \Delta x(E_0)/2. \quad (20)$$

The dependences of these quantities on E_0 and the xenon density are shown in Figs. 7 and 8. It should be noted that, generally, dependences (17)–(20) do not allow one to predict the plasma parameters in the cathode sheath. The matter is that the value of E_0 is not known *a priori* and, thus, should be measured. One only knows the voltage across the discharge gap, U_0, which is to be related to E_0. Formula (1) cannot be used for this purpose. Since the voltage drop $\Delta\varphi_0$ across the cathode sheath is low compared to U_0, we need to consider a region far from the cathode to find E_0, which is beyond the scope of this study.

3.2. Relaxation Times

The relaxation time of the electron density near the cathode is determined by the ionization frequency ν_i. At $E \sim 3 \times 10^4$ V/cm, this time is rather short: $\nu_i^{-1} \sim 5 \times 10^{-11}$ s at $N_{Xe} = 4.4 \times 10^{18}$ cm^{-3}.

The relaxation time of the ion current near the cathode ($x \ll \Delta x$), τ_{ion}, is significantly longer:

$$\tau_{ion} \sim (u_{di}(E)\Delta x)^{-1} \sim 10^{-7} \text{ s}.$$

Here, $u_{di}(E)$ is the ion drift velocity, which amounts to $\sim 5 \times 10^4$ cm/s at $E \sim 10^4$ V/cm. The optimum duration of the pumping pulse should be longer than τ_{ion}.

Between the pulses, recombination occurs. For the plasma in the cathode sheath not to recombine completely between the pulses, the repetition rate of the pumping pulses should not be too low.

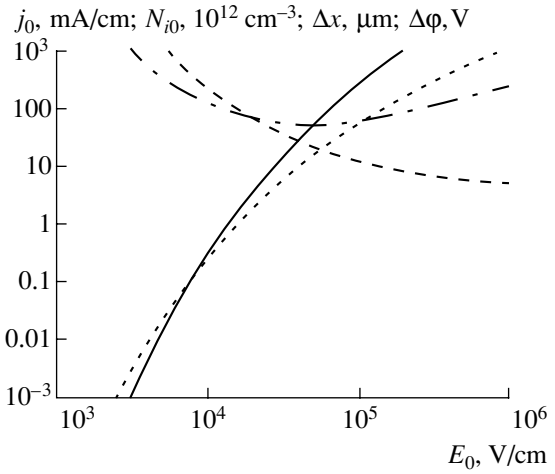


Fig. 7. Linear current density j_0 (solid curve), ion density N_{i0} (dotted curve), cathode sheath thickness Δx (dashed curve), and potential drop $\Delta\varphi_0$ (dashed-and-dotted curve) vs. electric field on the cathode E_0 for $N_{Xe} = 4.4 \times 10^{18}$ cm^{-3}.

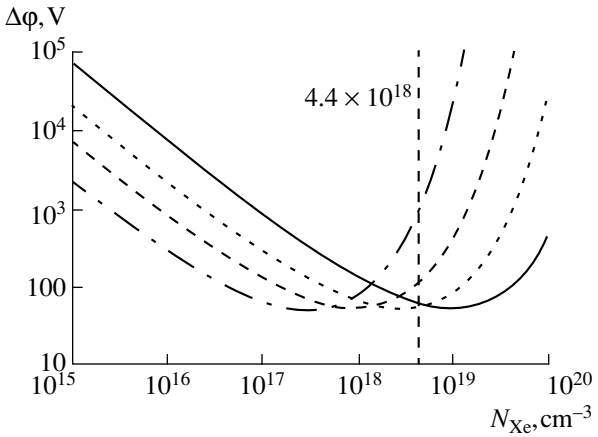


Fig. 8. Potential drop $\Delta\varphi_0$ vs. gas density for different electric fields at the cathode: $E_0 = 10^5$ (solid curve), 3×10^4 (dotted curve), 10^4 (dashed curve), and 3×10^3 V/cm (dashed-and-dotted curve).

Let us consider two recombination mechanisms for electrons and ions: the surface and the volume ones. On the cathode surface, the recombination time between the pumping pulses is determined by the ion diffusion time

$$\tau_{dif} \sim (v_{Xe}/\sigma_{tr} N_{Xe} \Delta x^2)^{-1} \sim 1.7 \times 10^{-5} \text{ s}$$

in a layer of thickness $\Delta x \sim 30$ μm. Here, $\sigma_{tr} \sim 1.2 \times 10^{14}$ cm^2 is the cross section for resonant charge transfer between Xe ions and atoms and $v_{Xe} \sim 2.7 \times 10^4$ cm/s is the thermal velocity of Xe atoms.

The characteristic time of volume recombination,

$$\tau_{dis} \sim (\alpha_{dis} N_i)^{-1} \sim 5 \times 10^{-7} \text{ s}$$

is determined by the dissociative recombination rate constant $\alpha_{dis} \sim 2 \times 10^{-6}$ cm^3/s, because at the pressures used, the conversion of atomic ions into molecular ions proceeds rather rapidly. Here, based on the calculated parameters of the cathode sheath, the ion density is set at $N_i \sim 10^{12}$ cm^{-3}.

It is seen that, at the pulse repetition rate in the range $f \sim$ 10–100 kHz, the ion density in the cathode sheath does not decrease below $N_i \sim 10^{10}$–10^{11} cm^{-3}.

4. CONCLUSION

In this study, we have simulated the development of an electron avalanche in xenon. Based on the simulation results and the available experimental data, analytical approximations of the quantities characterizing the electron multiplication in the cathode sheath have been proposed. The critical electric field has been found, above which the runaway of electrons occurs. It is shown that, in high-pressure xenon lamps, there are no runaway electrons; consequently, these electrons cannot be the main cause for the formation of a volume discharge in a single-barrier xenon excimer lamp.

Based on the approximations proposed, the dependences of the plasma parameters (the electron and ion densities and currents and the electric field) on the distance from cathode have been analyzed. Simple formulas that determine the total current, the cathode sheath thickness, and the voltage drop across the sheath as functions of the electric field on the cathode surface have been proposed.

The study above allows us to draw conclusions regarding the cathode sheath in a xenon lamp with a cylindrical electrode about one millimeter in diameter.

In the cathode sheath, electron multiplication occurs and the electric field is screened by the ion charge. At a pressure of ~100 torr and an electric field of about 10 kV/cm, the cathode sheath thickness turns out to be smaller than the wire radius by one order of magnitude. Because of the small sheath thickness, the voltage drop across the cathode sheath is significantly lower than the voltage applied to the discharge gap. At a given electric field at the cathode, the screening of the electric field in the cathode sheath limits the current density through the sheath.

The results obtained allow one to self-consistently consider the problem of the excitation of discharge xenon lamps.

ACKNOWLEDGMENTS

We are grateful to V.F. Tarasenko for stimulating discussions. This study was supported by the International Science and Technology Center, project no. 1270.

REFERENCES

1. U. Kogelshatz, Pure Appl. Chem. **62**, 1667 (1990).
2. W. Sasaki, S. Kudobera, and J. Kawanaka, Proc. SPIE **3092**, 378 (1997).
3. A. N. Panchenko, E. A. Sosnin, and V. F. Tarasenko, Opt. Commun. **161**, 249 (1990).
4. E. Arnold, R. Dreiskemper, and S. Reber, in *Proceedings of 8th International Symposium on Science and Technology of Light Sources, Graifswald, 1998*, p. 90.
5. F. Vollkommer and L. Hitzschke, in *Proceedings of 8th International Symposium on Science and Technology of Light Sources, Greifswald, 1998*, p. 51.
6. C. K. Rhodes and P. W. Hoff, in *Excimer Lasers*, Ed. by C. K. Rhodes (Springer-Verlag, New York, 1979; Mir, Moscow, 1981), Chap. 6.
7. C. Duzy and J. Boness, IEEE J. Quantum Electron. **16**, 640 (1980).
8. D. J. Eckstrom, H. H. Nakano, J. E. Lorents, *et al.*, J. Appl. Phys. **64**, 1679 (1988).
9. A. M. Boichenko, S. I. Yakovlenko, and V. F. Tarasenko, Laser Part. Beams **18**, 655 (2000).
10. S. I. Yakovlenko, in *Encyclopedia of Low Temperature Plasma*, Ed. by V. E. Fortov (Nauka, Moscow, 2000), Introductory Vol. 4, p. 262.
11. M. Surendra, D. B. Graves, and G. M. Jellum, Phys. Rev. A **41**, 1112 (1990).
12. *Physical Quantities: Handbook*, Ed. by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991).
13. Yu. P. Raizer, *Gas Discharge Physics* (Nauka, Moscow, 1987; Springer-Verlag, Berlin, 1991).
14. E. Krishnakumar and S. K. Srivastava, J. Phys. B **21**, 1055 (1998).
15. B. M. Smirnov, *Atomic Collisions and Elementary Processes in Plasma* (Atomizdat, Moscow, 1968).
16. A. M. Eletskiĭ and B. M. Smirnov, *Physical Processes in Gas Lasers* (Énergoatomizdat, Moscow, 1985).
17. S. I. Braginskii, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Gosatomizdat, Moscow, 1963; Consultants Bureau, New York, 1965), Vol. 1.
18. H. W. Drawin and F. Emard, Physica C (Amsterdam) **87**, 333 (1977).

Translated by N. Ustinovskiĭ

Technical Physics, Vol. 48, No. 2, 2003, pp. 199–204. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 65–70.

GAS DISCHARGES, PLASMA

Increase in the Power of Arc Cesium–Oxygen Thermionic Converters with Tungsten Electrodes at Elevated Emitter Temperatures

V. P. Kobyakov* and A. G. Kalandarishvili**

** Institute of Structural Macrokinetics and Materials Science, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
e-mail: koba@ism.ac.ru

*** Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia*
e-mail: kalandarishvili@dhtp.kiae.ru

Received July 15, 2002

Abstract—The properties of experimental thermionic converters with emitters made of oxygen-containing single-crystal tungsten are studied. The tungsten is produced by the chloride CVD method. Such emitters not only perform their basic function in the converter but also serve as an "internal" oxygen source. An attempt is made to raise the emitter temperature in order to increase the maximal output of the converter using the new emitter material. The efficiency of applying an oxygen-free single-crystal tungsten collector is contrasted with that when the collector is made of polycrystalline tungsten or Cb-1 niobium alloy. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Today, it is obvious that arc cesium thermionic converters (TCs) with single-crystal tungsten electrodes hold much promise [1]. However, a number of issues, such as the introduction of oxygen into the interelectrode space to improve the TC output and the choice of the most efficient collector material, are still incompletely understood. It was suggested [2] that oxygen be introduced from an internal source, single-crystal oxygen-containing tungsten [3–5]. In this work, we use this new emitter material with a "high-temperature" work function of 5.7 eV in pilot TCs.

Along with conventional Cb-1 niobium alloy, which contains about 1% Zr, several oxidized refractory metals have been studied [6–11] in an attempt to improve TC output. However, as early as about 40 years ago, Bol'shov and Zharinov [12] noted that the collector surface in an operating TC will inevitably be contaminated by emitter material evaporation products. Therefore, later [13] TCs with a polycrystalline tungsten collector were studied. In this work, we apply a collector made of single-crystal tungsten and compare the associated TC with those where the collector is made of polycrystalline tungsten or Cb-1 alloy. Earlier [14], the single-crystal tungsten collector was used in experimental TCs with cesium–barium fillers in an effort to achieve the maximal adsorption of barium.

Of interest also is the examination of the output characteristics of cesium–oxygen TCs with tungsten electrodes at elevated temperatures. Such a seemingly easy way to improve TC output runs into difficulties in terms of materials science. Estimates and experiments made for an oxygen-free arc cesium TC of cylindrical geometry (where the emitter was made of zirconium carbide and the collector, of aluminum) have shown [15] that, with an electrode spacing of about 5 mm, the power density at a level $P \geq 40$ W/cm^2 can be provided at emitter temperatures as high as $T_e \geq 2500$ K. Under these conditions, the efficiency has been demonstrated to exceed 16%, reaching 25% at $T_e = 3050$ K and $P = 106$ W/cm^2. Our efforts to approach these temperatures in this work have not met with much success.

The goal of this study was to gain additional information concerning the three aspects mentioned above.

EXPERIMENTAL

The studies were performed on experimental cylindrical monomodular TCs. Their emitters were made by the original CVD technique [3], which provides the production of tubular tungsten single crystals faceted by six (110) planes and containing an increased amount of oxygen (10^{-3}–10^{-2} wt %) in the form of a quasi-solid solution. Such emitters exhibit a sharp temperature dependence of the work function, which reaches 5.6–5.7 eV at temperatures above 2000 K [2, 5]. Before being used in cylindrical TCs, single-crystal tungsten tubes were turned with a lathe and then electrochemically polished in an alkaline solution to remove the damaged layer (about 100 μm thick). After such processing, the work function declined to 5.2 eV. Tubular rear parts, one of which had a dead end, were welded to

1063-7842/03/4802-0199$24.00

both sides of the emitter by electron beam welding. The rear parts had special positioning shoulders, which were seated against the ceramic belts of the collector unit upon assembling and thus specified the electrode spacing. A tubular tungsten resistance-type heater was inserted into the emitter. A hole to receive a W(5% Re)/W(20% Re) tungsten–rhenium thermocouple was predrilled on the emitter end face.

The collector unit was prepared by the gas-static pressing of a three-layer tubular pile [16, 17]. After the pressing, the inner metal layer, which served as a coat through which the ceramic layer was consolidated and simultaneously the collector and ceramic layer were diffusion-welded to the outer metal layer, was removed by turning. With this coat removed, the ceramic and collector layers became exposed and the collector layer was electrochemically finished. Thus, the inner layer of the collector bank was composite. After assembling, the central part of the collector unit, where the collector itself is located, faced the emitter. The remaining part of the collector unit had an inner insulating ceramic coating. As a collector layer, we used tubular single-crystal tungsten with the ⟨111⟩ axis grown by oxygen-free chloride CVD. For comparison, collectors were made of polycrystalline tungsten and Cb-1 niobium alloy (1 wt % Zn + 0.1 wt % C). A branch pipe for evacuating the module and for cesium vapor delivery was welded into the side of the collector unit, and a tubular vacuum-tight Nb–Al_2O_3–Nb pressure seal was welded to the open end of the module.

Upon assembling the module, the dead end of the emitter unit with the calked-in W(5% Re)/W(20% Re) thermocouple was inserted into the collector cavity, which is also dead at one end. The rare part of the open end of the emitter unit had an expansion whose diameter was matched to that of the pressure seal welded to the collector. Once the emitter unit had been completely mounted, the emitter and collector faced each other, the expanded end was in intimate contact with the cylindrical niobium gland of the seal, and the emitter–collector spacing (0.25–0.30 mm in our case) was specified by the positioning shoulders. The assembling procedure was complete when the end face of the gland was circumferentially soldered to the expansion of the emitter rare part by electron-beam welding.

The assembly (module) was placed on a test bench to measure its output characteristics. The module was insulated from the bench by ceramic sleeves. The working section of the bench consists of three hollow coaxial cylinders producing two isolated cylindrical spaces. The inner space, providing gas cooling, was in direct contact with the collector. The outer space was a water jacket. By varying the pressure in the gas space, as well as the water flow rate in the jacket, one could vary conditions for heat transfer from the module in operation to the coolant and thus control the collector temperature over wide limits. Appropriate temperature conditions during the tests were established with electrical heaters (a nichrome wire in a quartz sheath) mounted on the outer surface of the working section. The power of the heaters was set and maintained with current- and voltage-controlling VRT-3 temperature regulators. The temperature was measured with Chromel–Alumel thermocouples mounted on the outer surface of the working section and on the module. The characteristic working parameters were T = 773–1073 K, pressure in the gas space P = 0.7–700 Pa, and water flow rate in the jacket G = 0.5 l/s.

The vacuum–cesium system (VCS) provided the desired pressure in the interelectrode space before cesium vapor delivery, the maintenance and control of the cesium vapor pressure within given limits, and the cesium condensation at a given site at the final test stage. The VCS basic element is a two-chamber cesium thermostat. One chamber is filled with cesium, while in the other, cooling tap water circulates. The cesium chamber is connected to the interelectrode space of the module and to high- and low-pressure vacuum pumps. The thermostat is communicated with the interelectrode space and pumps (in turn or simultaneously) via a tee. The thermostat and the VCS communications throughout their length are equipped with electrical heaters controlled by the VRT-3 temperature regulators.

The residual pressure in the interelectrode space was measured with a VIT-3 vacuum gauge. The test conditions provided a static residual pressure in the interelectrode space of no more than $(1–3) \times 10^{-4}$ Pa. The cesium vapor pressure in the interelectrode space was varied by varying the thermostat temperature and was kept between 300 and 700 Pa. The temperature of the thermostat and pipelines was measured with the Chromel–Alumel thermocouples.

Once the whole system had been degassed and the module had operated as a converter for about 100 h, we took the current–voltage characteristics $j = f(U)$ at different emitter (T_e) and collector (T_c) temperatures. Then, the heaters of the VCS were switched off. As the thermostat temperature was reduced to T = 320–350 K, water started circulating in the cooler of the thermostat, providing the condensation of the cesium vapor in it.

RESULTS

Figure 1 shows examples of the current–voltage curves and associated dependences of the maximal output power density on the output voltage, $P = f(U)$, for the module with the single-crystal tungsten electrodes at T_e = 2100 K and T_c = 800–1300 K. From the current–voltage characteristics, the dependences $P = f(T_e)$ and $P = f(T_c)$ (Figs. 2, 3) for the modules with different collectors were constructed.

From Fig. 2, it follows that, first, the output of the modules with the tungsten collectors is more than twice as high as that for the module where the collector is made of the niobium alloy (cf. curves *a*, *b*, and *c*). Second, the efficiency of the tungsten collector, unlike the

emitter, is almost independent of whether single-crystal or polycrystalline tungsten is used (cf. curves *a* and *b*). Third, the output vs. emitter temperature dependence has two clear-cut portions in all the cases. At temperatures up to 1900–1950 K, the curves rise only slightly or even tend to saturate. At higher temperatures, the output grows sharply. In the case of the niobium collector, the rise rapidly changes to the tendency toward saturation; for the tungsten collectors, the curves continue to increase sharply. Unfortunately, we could not increase the temperature above 2100 K with our bench; however, the run of these curves implies that they do not even approach saturation.

All the dependences $P = f(T_e)$ in Fig. 3 have a peak. The peaks in the panel *c* (niobium collector) appear at a collector temperature of 900 K. For the modules with tungsten collectors (panels *a*, *b*), the peak corresponds to collector temperatures of no lower than 1000 K.

DISCUSSION

As was mentioned in the Introduction, one of the main goals of our study was to estimate the efficiency of oxygen-containing single-crystal CVD tungsten as an emitter material of cesium–oxygen TCs [3]. This emitter material offers high adsorptivity and emissivity and additionally can serve as a source of oxygen arriving from the bulk of the emitter directly at the emitting surface as the emitter is heated [4]. Experimental results reported in the previous section (Figs. 1–3) demonstrate various aspects of the oxygen effect. Impressive are the sharply increasing temperature dependences of the power generated at emitter temperatures higher than 1900–1950 K for the modules with the tungsten collectors (Fig. 2). It was shown [4, 5] that at temperatures above 1900 K, the oxygen flow from the bulk to the surface is greatly enhanced, substantially increasing the work function of this surface. With the cesium vapor present over the emitter surface, cesium is adsorbed on the surface more intensely. It was demonstrated [4] that the temperature dependence of the output of experimental TCs with emitters made of oxygen-containing single-crystal tungsten is well approximated by a straight line when plotted in the Arrhenius coordinates. The apparent energy of activation found from the slope of this line turned out to be close to that of gaseous impurity diffusion in tungsten.

Figure 4 shows the Arrhenius plot of the output constructed from the averaged data for the modules with oxygen-containing tungsten emitters. Data points are seen to lie well in a straight line (a correlation coefficient $r = 0.984$), which allows us to extrapolate it. The output was found to be zero at $T = 1458$ K, i.e., precisely in the middle of the interval 1400–1500 K, where noticeable electron emission from the tungsten surface is usually observed. Since the emitter temperature was limited by 2100 K in our tests, extrapolation toward higher temperatures was also of interest. As follows from Fig. 4, the output extrapolated to $T_e = 2500$ K reaches 80 W/cm^2.

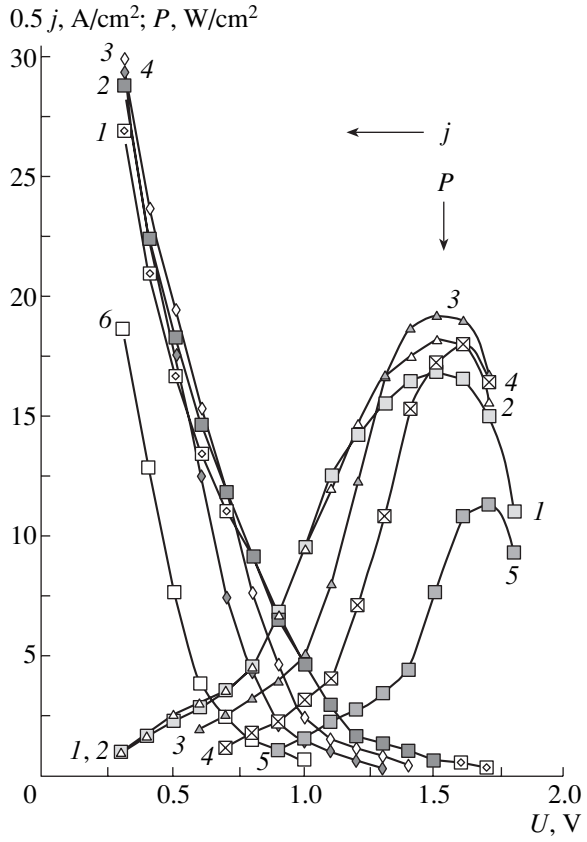


Fig. 1. Current–voltage characteristics of an experimental arc cesium–oxygen TC with single-crystal tungsten electrodes. The emitter temperature is 2100 K. The collector temperature is (*1*) 800, (*2*) 900, (*3*) 1000, (*4*) 1100, (*5*) 1200, and (*6*) 1300 K.

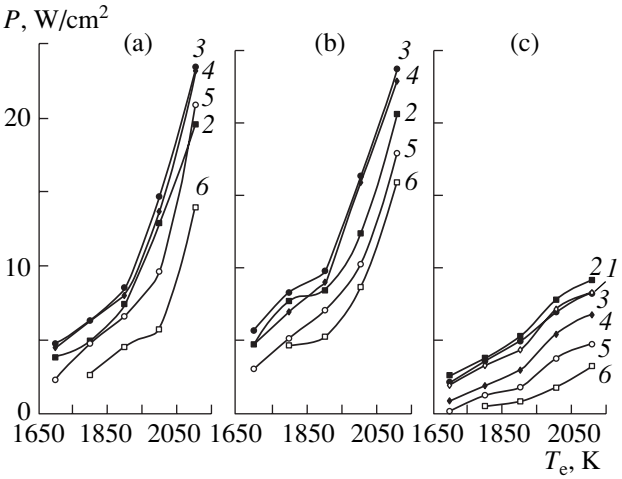


Fig. 2. Output vs. emitter temperature dependence for modules with different collector materials: (a) single-crystal W⟨111⟩, (b) polycrystalline W, and (c) Cb-1 alloy. The collector temperatures are (*1*) 800, (*2*) 900, (*3*) 1000, (*4*) 1100, (*5*) 1200, and (*6*) 1300 K.

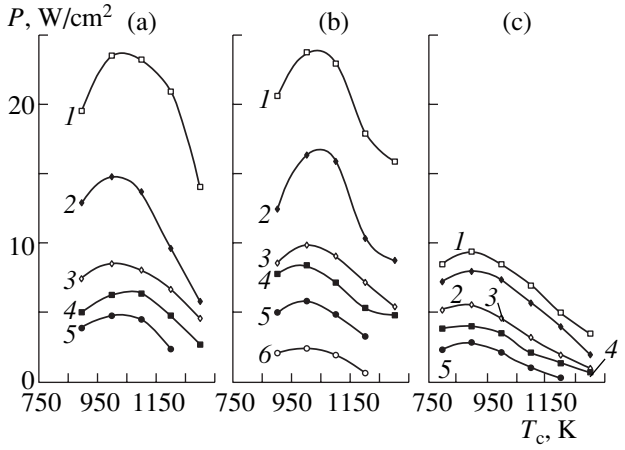


Fig. 3. Output vs. collector temperature dependence for modules with different collector materials: (a) single-crystal W⟨111⟩, (b) polycrystalline W, and (c) Cb-1 alloy. The emitter temperatures are (*1*) 2100, (*2*) 2000, (*3*) 1900, (*4*) 1800, (*5*) 1700, and (*6*) 1600 K.

The results of this work was compared with those found earlier for experimental cesium–oxygen and high-temperature TCs (Fig. 5). The most complete data were published 30 years ago. In [18], cylindrical modules with an emitter made of tungsten produced by chloride CVD and ⟨110⟩ textured were studied. Unlike the tungsten used in our work, that in [18] is not single-crystal and, moreover, is oxygen-free (the oxygen content in this material is at least no higher than 10^{-5} wt %). The collector was covered by an oxygen-containing molybdenum coating applied by vacuum evaporation. The output averaged over five modules was found to be 9 W/cm^2. One of the most carefully designed modules (unfortunately with plane-parallel electrodes) was analyzed in [7]. The emitter material, tungsten with the ⟨110⟩ texture obtained by chloride CVD, had a work function of 4.85 eV. The collector was made of nontextured polycrystalline tungsten. The emitter was electrochemically polished and decarbonized by heating under a low oxygen pressure at T_e = 1900 K. Then, the cesium vapor was delivered, the current–voltage characteristics were taken, and the cesium was removed. The inner surfaces of the module were oxidized at T_e = 1073 K, and the work function as a function of T_c was measured. As T_c grew, so did the emitter current, which is a direct indication that oxygen comes to the emitter surface from the collector. The current–voltage characteristic taken after the oxidation showed a much higher output than before the oxidation: at the same current density

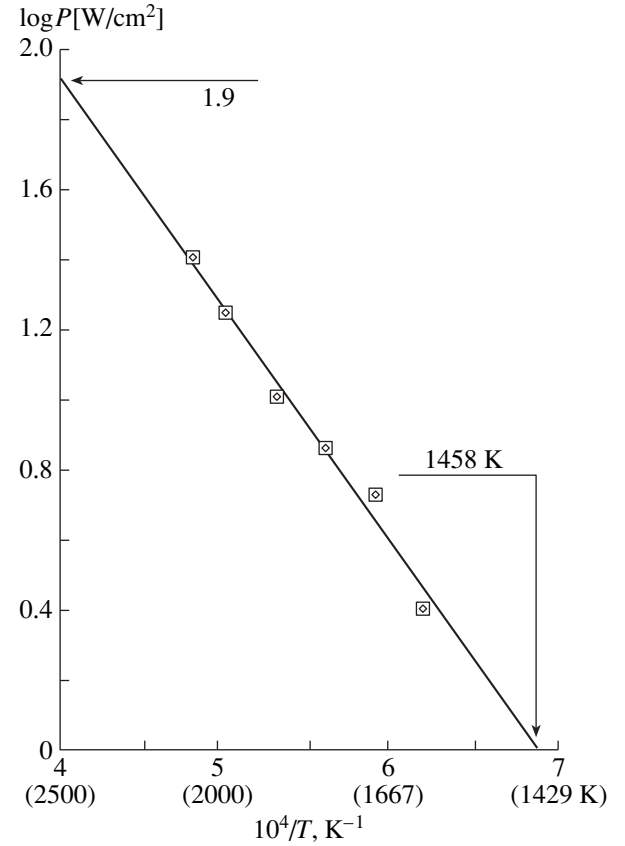


Fig. 4. Polytherm (in the Arrhenius coordinates) of the output of oxygen-containing modules (averaged over data presented in this work). The arrows indicate extrapolated values.

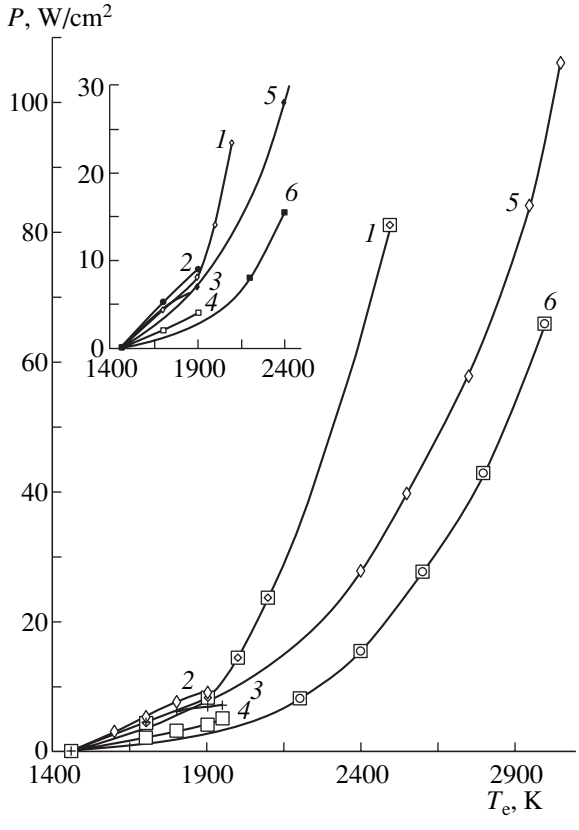


Fig. 5. Polytherm of the output of oxygen-containing modules: (*1*) our data for the module with the single-crystal tungsten collector (the top-most point is found by extrapolation), (*2*) [6], (*3*, *4*) [7], (*5*) [15], and (*6*) [21]. All the curves are extrapolated to the zero power.

10 A/cm^2, the output voltage increased more than twice from 0.2 to 0.46 V. The output vs. T_e dependences before and after the oxidation were also given (curves *3* and *4* in Fig. 5). Note that the long-term testing of the cesium–oxygen module performed in [7] showed that the output was stable within several percent during the operation for 1400 h.

The results cited above and our data (curve *1*) indicate that, at temperatures between 1500 and 1900 K, the output of all the cesium–oxygen TCs is roughly the same and twice as high as that of the TCs where the amount of oxygen is insufficient for the oxygen effect to be significant (cf. curves in Fig. 2a–2c and curves *3* and *4* in Fig. 5). Unfortunately, we failed to find published data for the high-temperature operation of cesium–oxygen TCs. As for our results (curve *1* in Fig. 5), they show that the output rises sharply at 1900–1950 K. However, for high-temperature oxygen-free modules [15, 19], the rise is not so dramatic (curves *5* and *6* in Fig. 5). It is seen that at T = 2500 K, the characteristics of these modules are on average thrice as low as in our case. This difference seems to increase still further with temperature, as follows from the run of the associated curves.

Thus, oxygen-containing single-crystal tungsten is a promising emitter material for arc cesium–oxygen TCs, its advantage being exploited most fully at high temperatures.

Some points on the choice of the emitter material need to be made. The general considerations concerning the steady-state condition of the operating TC surface are not quite adequate. When discussing Fig. 2, we noted that the difference between the curves $P = f(T_e)$ for the niobium and tungsten collectors is that the former not only runs below the latter but also tends to saturate at temperatures higher than 2000 K. It should be emphasized here that the emitter material in all the TC experimental designs was the same: oxygen-containing single-crystal tungsten. Therefore, the only reason for this difference is the difference in the collector material. It can thus be concluded that the use of tungsten collectors in cesium-oxygen TCs is particularly promising in the sense that they do not absorb the oxygen entering into the interelectrode space from the emitter. In this sense, it really does not matter whether single-crystal or polycrystalline tungsten is applied. The same emitter, when combined with zirconium-doped niobium collector, is half as efficient, since the oxygen from the gas phase is absorbed by niobium and then irreversibly combines with the zirconium present in the alloy to form zirconium oxide [20]. Eventually, the steady state, which is characterized by the saturation of the dependences $P = f(T_e)$, is established (Fig. 2).

Thus, materials that do not absorb oxygen, such as tungsten, are the materials of choice for cesium–oxygen TCs.

Of great interest are the curves $P = f(T_c)$ in Fig. 3. It was noted in the previous section that the optimal temperature of tungsten collectors in cesium–oxygen TCs can be substantially increased. For oxygen-free modules with niobium collectors, the optimal collector temperature is near 900 K (curve in Fig. 3c). At 1100 K, the output drops to 7 W/cm^2. Temperature optimum in this case reflects the establishment of the complex steady state influenced by many factors. For oxygen-containing modules with tungsten electrodes, the optimum shifts to 1000–1100 K (curves in Fig. 3a and 3b). With the collector temperature increased by further 200 K, the output remains at a level of 18–20 W/cm^2. This still more complex steady state is affected by additional factors due to the presence of oxygen in the interelectrode gap. For detailed analysis of this effect, additional experimental data (omitted in this paper) must be invoked.

Since actual nuclear power plants with a TC [1] are designed for operation in space, heat removal from the collector is accomplished by radiation alone; that is, a heat sink is a basic component of a TC. An increase in the radiator temperature by 200–300 K would allow engineers to considerably reduce the weight and overall dimensions of a heat sink and the system as whole.

CONCLUSIONS

(1) Oxygen-containing single-crystal tungsten is a promising material for emitters of cesium–oxygen TCs.

(2) The maximum efficiency of this emitter material is reached at temperatures of higher than 1900 K.

(3) The maximum efficiency of cesium–oxygen TCs can be reached by using collector materials that do not absorb oxygen.

(4) When tungsten is used as a collector material that does not absorb oxygen, the optimal collector temperature can be raised to 1000–1100 K and, accordingly, the weight and overall dimensions of related power plants can be greatly reduced.

ACKNOWLEDGMENTS

The authors are indebted to Yu.L. Éloshvili, B.I. Ermilov, and V.G. Kashiya for the fabrication and testing of experimental TC modules.

REFERENCES

1. G. M. Gryaznov, E. E. Zhabotinskiĭ, A. V. Zrodnikov, *et al.*, At. Energ. **66**, 374 (1989).
2. V. P. Kobyakov, Zh. Tekh. Fiz. **66** (7), 169 (1996) [Tech. Phys. **41**, 726 (1996)].
3. V. P. Kobyakov, Kristallografiya **41**, 552 (1996) [Crystallogr. Rep. **41**, 523 (1996)].
4. V. P. Kobyakov, Zh. Tekh. Fiz. **66** (7), 161 (1996) [Tech. Phys. **41**, 722 (1996)].
5. V. P. Kobyakov, V. A. Koryukin, and V. P. Obrezumov, Zh. Tekh. Fiz. **66** (8), 169 (1996) [Tech. Phys. **41**, 834 (1996)].

6. D. M. Ernst, in *Proceedings of IX IEEE Conference on Rec. Thermion. Convers. Specialist* (New York, 1970), p. 492.
7. D. Lieb and F. Rufeh, in *Proceedings of 3rd International Conference on Thermion. Electron Power Generation* (FRG, 1972), Vol. 3, p. 1091.
8. D. Lieb, D. Goodale, and T. Briere, in *Proceedings of 12th Intersociety Energy Conv. Engin. Conference* (Washington, 1977), p. 1555.
9. V. K. Tskhakaya and V. I. Yarygin, *Thermoemission Transformation of Thermal Energy into Electric Energy* (FÉI, Obninsk, 1980), p. 136.
10. N. E. Menabde and L. M. Tsakadze, in *Proceedings of 2nd Branch Conference on Nuclear Energetics in Space* (SFTI, Sukhumi, 1992), p. 81.
11. V. P. Kobyakov and A. G. Kalandarishvili, Zh. Tekh. Fiz. **68** (8), 131 (1998) [Tech. Phys. **43**, 997 (1998)].
12. V. G. Bol'shov and A. A. Zharinov, Zh. Tekh. Fiz. **32** (2), 214 (1962) [Sov. Phys. Tech. Phys. **7**, 152 (1962)].
13. F. Morris, in *Proceedings of 12th Intersociety Energy Conv. Engin. Conference* (Washington, 1977), p. 1540.
14. V. D. Bondarenko, M. B. Kraĭz, A. I. Loshkarev, *et al.*, in *Proceedings of XIV All-Union Conference on Emission Electronics* (Fan, Tashkent, 1970), Sect. 1, p. 83.
15. T. V. Krachino and T. L. Matskevich, Zh. Tekh. Fiz. **38**, 1395 (1968) [Sov. Phys. Tech. Phys. **13**, 1140 (1968)].
16. C. V. Weaver and W. A. Ranken, in *Proceedings of Thermion. Convers. Specialist Conference* (Carmel, New York, 1969), p. 347.
17. P. Fiebelmann, Forsch. Geb. Ingenieurwes **38** (5), 133 (1972).
18. D. M. Ernst, in *Proceedings of IEEE Conference on Rec. Thermion. Convers. Specialist, 9th Annual Conference, Miami Beach* (New York, 1970), p. 492.
19. T. L. Matskevich and T. V. Krachino, Zh. Tekh. Fiz. **41**, 1182 (1971) [Sov. Phys. Tech. Phys. **16**, 934 (1971)].
20. V. Ch. Gonchikov, A. D. Korotaev, A. N. Tyumentsev, *et al.*, in *High-Purity and Monocrystalline Metal Materials*, Ed. by A. I. Manokhin (Nauka, Moscow, 1987), pp. 161–165.

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 205–208. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 71–74.

GAS DISCHARGES, PLASMA

Spatial Structure of the Lower Mode of Ion-Acoustic Instability in the Plasma of a High-Current Ion Laser

V. I. Donin*, V. A. Ivanov*, V. V. Pikalov, and D. V. Yakovin***

** Institute of Automation and Electrometry, Siberian Division, Russian Academy of Sciences, Universitetskiĭ pr. 1, Novosibirsk, 630090 Russia*
e-mail: donin@iae.nsk.su

*** Institute of Theoretical and Applied Mechanics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

Received April 12, 2002

Abstract—The spatial structure of the lower mode of ion-acoustic instability in the plasma of the continuous discharge of a high-current argon ion laser is studied experimentally. The oscillations in the intensity of the spontaneous ArII line radiation from the plasma are recorded in different directions in cross sections perpendicular to the discharge axis. The power spectra of the integral projections of radiation are obtained and used to reconstruct the two-dimensional spatial structure of the mode, which is identified by using the available theoretical models. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

At present, ion gas lasers are the most powerful sources of continuous coherent radiation in the visible and near ultraviolet spectral regions. As an active medium of these lasers, low-pressure high-current discharges generated in cylindrical tubes $5 \le d \le 30$ mm in diameter are used [1]. Earlier, it was experimentally established that, under optimal (with respect to the generation power) conditions, in addition to high-frequency ion-acoustic oscillations at frequencies of $\omega \lesssim \Omega_i$ (where Ω_i is the ion plasma frequency), low-frequency ion-acoustic oscillations at frequencies of $\omega \lesssim 10^{-2}\Omega_i$ [1–5] can spontaneously be excited in the plasma of an ion laser. The threshold current for the onset of instability depends on d, the gas pressure, and the distribution of atoms along the discharge. The increase in the current above the threshold value can result in the destruction of the discharge tube wall, which limits the output power and the laser service life [1].

The low-frequency spectrum of ion-acoustic instability in an argon laser plasma consists of individual peaks within the range $0.1 \lesssim \nu \lesssim 2$ MHz (where $\nu = \omega/2\pi$). Near the threshold, there are no more than two peaks. The number of peaks increases with increasing discharge current. These low-frequency peaks are the lower modes of long-wavelength ion-acoustic oscillations of the plasma bounded by the tube wall [6, 7]. The experimentally determined dispersion relation of the lower (with the lowest frequency) mode [8] is in good agreement with the theory of ion-acoustic oscillations in low-pressure cylindrical discharges [9, 10]. To our knowledge, the spatial inhomogeneity of the oscillation intensity of the two lowest modes was experimentally studied only in [7]. The qualitative radial dependence of the oscillation intensity obtained in [7] was insufficient to provide an idea about the spatial structure of the modes and the number of each mode. In the present paper, the two-dimension spatial structure of the lowest frequency mode is determined by the tomographic method, and this mode is identified according to the available theoretical models.

EXPERIMENTAL SETUP AND RESULTS

An argon plasma was produced in a steady-state high-current discharge in a tube 1 m in length and 16 mm in diameter (Fig. 1). The tube consisted of water-cooled aluminum sections with an oxide coating

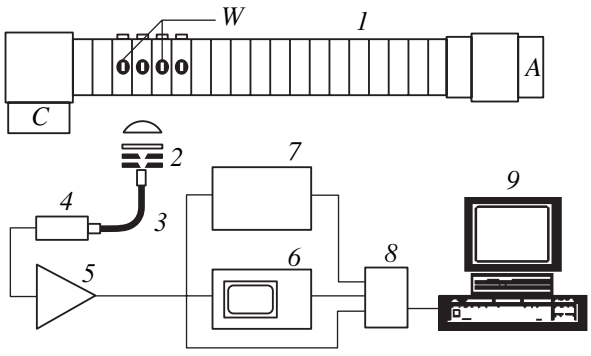


Fig. 1. Block diagram of the experimental setup for studying the spatial structure of plasma oscillations: (*1*) discharge tube with cathode (*C*), anode (*A*), and windows (*W*); (*2*) optical system with fiber (*3*); (*4*) photomultiplier; (*5*) amplifier; (*6*) spectrum analyzer; (*7*) correlator; (*8*) analog-to-digital converter; and (*9*) computer.

1063-7842/03/4802-0205$24.00 © 2003 MAIK "Nauka/Interperiodica"

and had a cool arc cathode with self-heated refractory bushing [1]. Four tube sections had openings made in the form of 4×16-mm slots located perpendicularly to the discharge axis (z axis) in order to output radiation from the discharge through quartz windows. Each section had three windows permitting observations at angles ξ of 0°, 55°, and 90° in the cross section perpendicular to the z axis. The sections with windows were located in the cathode region of the tube, where the oscillations under study were most intense (the corresponding intensity profile is given in [8]). At argon pressures of 0.2–0.4 torr (the discharge pressure was much lower, especially in the cathode region), the threshold current was in the range 300–400 A. The measurements were carried out under conditions close to the instability threshold. The lowest oscillation mode had a frequency of $\nu_0 = 190$ kHz.

To investigate the spatial structure of the oscillations, we used the method of plasma emission tomography [11]. The intensity of the plasma emission was determined mainly by blue-green lines of singly excited argon ions with the density n_i^*. At currents slightly exceeding the threshold current, the density of ions in the ground state is $n_i(t) = n_i + \delta n_i(t)$, where n_i is the time-independent component and $\delta n_i(t)$ is a small ($\lesssim$3% of n_i) oscillating component. Under the experimental conditions, the density n_i^* is close to saturation with respect to the electron density; i.e., $n_i^*(t) \sim n_i(t)$,

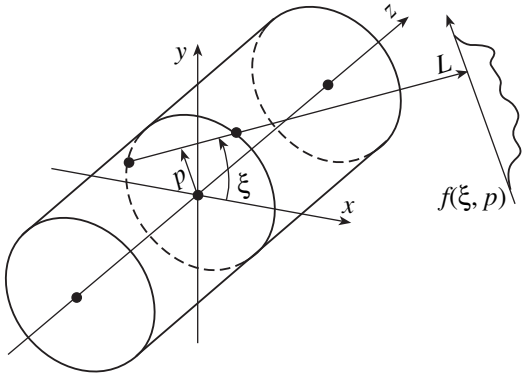


Fig. 2. Tomographic scheme for observing plasma emission.

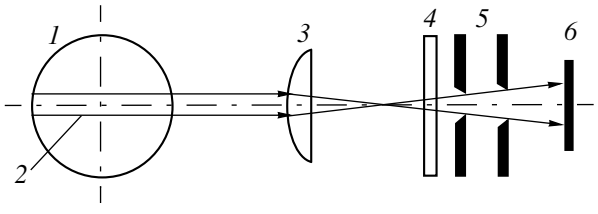


Fig. 3. Optical scheme for recording plasma emission: (*1*) discharge cross section, (*2*) cylindrical volume element, (*3*) lens, (*4*) light filter, (*5*) apertures, and (*6*) receiving fiber platform.

at least for frequencies of $\nu < \beta_i$, where β_i is the ionization rate [1].

The tomographic method consists in measuring the plasma emission at different angles in the cross sections perpendicular to the z axis. A scheme of the measurements is shown in Fig. 2, where x and y are the coordinates of points in the cross section under study. At each instant, the intensity of the plasma radiation emitted in the observation direction L is proportional to the integral of the density of the excited ions along the line of sight

$$\int_L n_i^*(x, y, t)dl,$$

where dl is the length increment along L.

In the experiment, the integral emission was recorded as one-dimensional projections $f(\xi, p)$ in the system of parallel rays specified by the angle ξ and the distance p from the z axis. Assuming that the plasma is optically thin, the projections are described by the classical Radon integral transform R [11]

$$f(\xi, p, t) = R\{n_i^*(x, y, t)\}. \quad (1)$$

From the projections experimentally measured at an arbitrary instant for the set of angles $\xi \in [0, \pi]$ and the set of coordinates $p \in [-d/2, d/2]$, one can obtain an estimate for the solution to integral equation (1) in the form

$$n_i(x, y, t) \sim R^{-1}\{f(\xi, p, t)\}, \quad (2)$$

where R^{-1} is the approximation of the inverse Radon transform.

We apply the Fourier transform in time to both sides of formula (2):

$$\tilde{n}(x, y, \nu) \sim R^{-1}\{\tilde{f}(\xi, p, \nu)\}. \quad (3)$$

Here and below, a tilde stands for the Fourier transform in time. The possibility of obtaining Fourier transform (3) of inverse-problem solution (2) was justified by numerical simulations. The frequency dependence of the function $\tilde{n}_i(x, y, \nu)$ at each point of the region under study represents the spectrum of the plasma ion oscillations. At a fixed mode frequency, this function represents its spatial distribution.

The projections and their Fourier spectra were measured in the experimental setup shown in Fig. 1. The plasma radiation emitted through a window was converted into an electrical signal with the help of an optical system consisting of a focusing lens, a blue-green light filter, two apertures, an optical fiber (Fig. 3), and a photomultiplier. Two identical systems were used in the measurements. Each system measured the plasma emission from a narrow cylindrical volume perpendicular to the discharge axis. The aperture diameters and the distances between the elements of the optical system were chosen according to the calculations of [12]. The spatial resolution of the optical system was 0.25 mm. Each system was situated on an individual

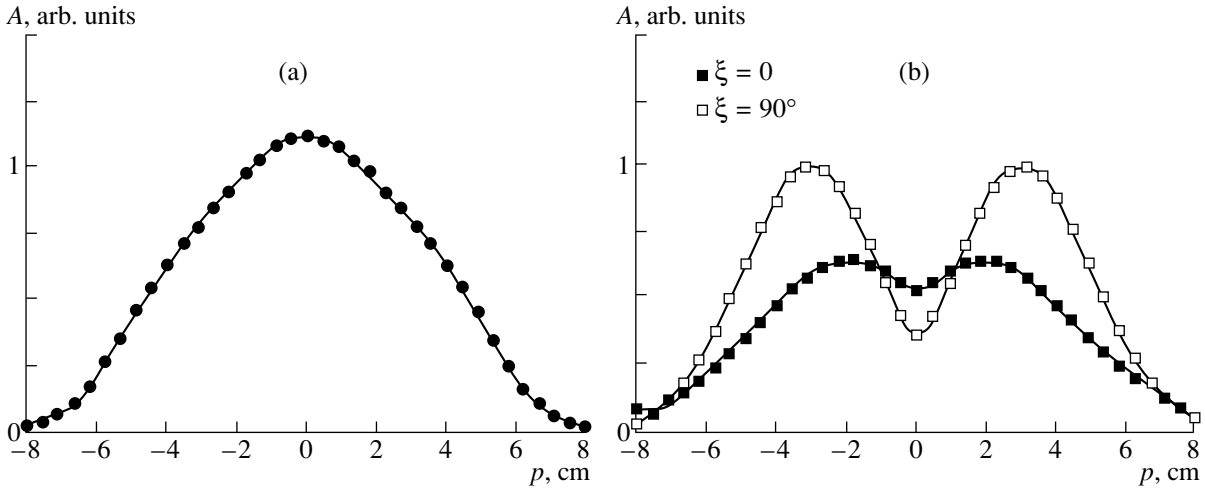


Fig. 4. (a) Time-independent and (b) oscillating components of integral plasma emission as functions of p.

bench, which could be smoothly displaced along the p coordinate perpendicular to the z axis with the help of an electric motor. One of these systems allowed consecutive measurements from two windows of the same section.

Since the available multichannel analog-to-digital converter (ADC) had a relatively low operation speed (≥30 μs), we used an SK4-59 analog spectrum analyzer; the voltage from the photomultiplier was fed at the input of the analyzer through an amplifier. The analyzer measured the amplitude of the harmonic component of the input signal at the frequency ν_0. The analyzer passband at this frequency was 10 kHz. The voltage reading from the analyzer output was proportional to the measured amplitude and represented the power spectrum of the signal $U \sim |\tilde{f}(\xi, p, \nu_0)|$. To determine the correlation characteristics of the oscillations (see [8]), the amplified signals from the photomultipliers were input to a Kh6-4 correlator. To record the correlation functions and the function $U(p)$, we used the ADC and a PC. The time-integrated signals from the amplifiers were fed to the inputs of the ADC. Each signal reflected the dependence of the constant component of the integral plasma emission on p in one of the observation directions (Fig. 4a). Using these curves, we could check the adjustment of the optical systems. Figure 4b shows typical power spectra for the projections at the frequency ν_0, measured simultaneously at two angles. The projections have a minimum at the center of the discharge between two equal maxima. Depending on the observation angle, the distance between the maxima could change (even up to the coincidence of the peaks). The measurements showed that, the displacement of the optical system along the z axis (within the 4-mm slit width) did not affect the shape of the projection. This means that the spatial distribution of the oscillations under study is almost uniform along the discharge axis; i.e., $|\mathbf{k}| \gg k_z$ where $\mathbf{k}$ is the wave vector of the oscillations and k_z is its longitudinal component.

The problem of reconstructing the spatial structure of the oscillation mode under study was to find its two-dimensional distribution by using the sets of the experimental projections by the method of computational tomography according to formula (3). In order to use formula (3), the imaginary and the real parts of the spectrum $\tilde{f}(\xi, p, \nu_0)$ must be separately measured experimentally (or calculated in one way or another). However the analyzer measured only the power spectrum of the projections. In the general case, the distribution reconstructed from the envelope of the projection spectrum may differ from the spatial structure of the function for which the projections were calculated. However, for a number of objects (including, as numerical simulations have shown, our case too) describing stationary harmonic oscillations, this method allows one to determine their characteristic spatial structure without substantially distorting it.

The tomograms were reconstructed using the method of maximum entropy [13], which is efficient for problems involving parallel measurements of the projections with a very small number of observation angles. In the calculations, we used the MENT algorithm included in the TOPAS–MICRO program package for the computational tomography of a plasma and gas [11]. The reconstructed tomograms (Fig. 5a) had two principal maxima that were symmetrical about the discharge center. The cross correlation functions of the oscillations measured from the different discharge regions showed the existence of a phase shift of π rad between the principal maxima. The artifacts in the tomograms (such as polygonal deformations) appeared as a result of a small number of observation angles. A comparison of the tomograms for the neighboring sections showed that the characteristic structures of the image turned by a small angle about the z axis. There-

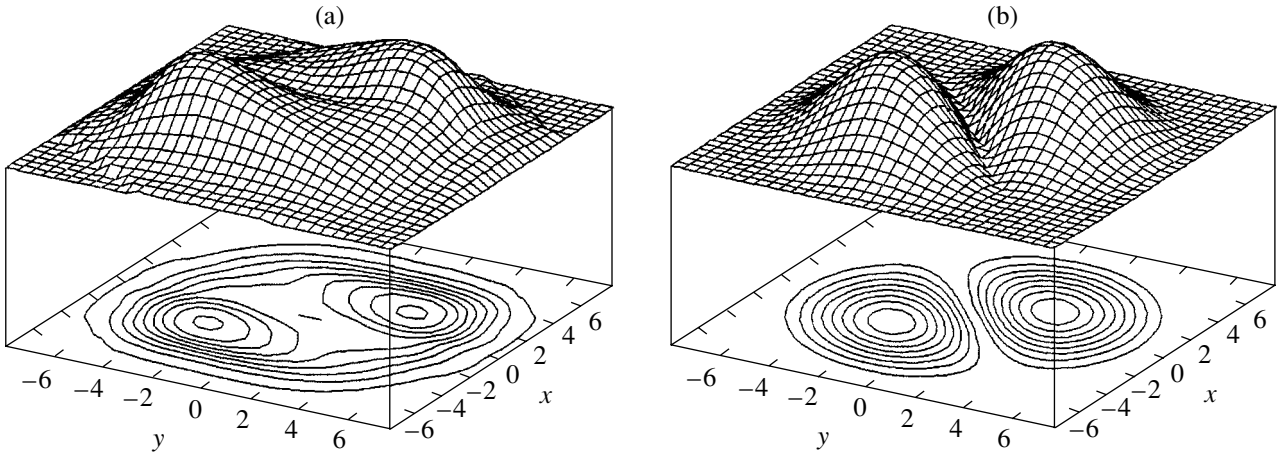


Fig. 5. (a) Reconstructed tomogram and (b) the structure (envelope) of the (1, 1) mode [7].

fore, the three-dimensional structure of the mode under study could qualitatively be represented as a double spiral with a turn period (according to our estimates) of 50–100 cm. The reasons why the mode structure turns slightly as z changes were not examined.

DISCUSSION OF THE RESULTS AND CONCLUSIONS

As was stated above, the spatial inhomogeneity of the intensity distribution of ion-acoustic oscillations in the high-current discharge of an ion laser was revealed experimentally in [7] from the qualitative difference of the radial distributions of the two lowest oscillation modes. In contrast to the experimental scheme used in that paper, our scheme allowed us to investigate the spatial structure of the oscillations by the tomographic method with a much better spatial resolution. The two-dimensional structure of the lowest mode was determined from the measurements of the power spectra of the integral projections and their processing by the method of computational tomography. The correlation measurements revealed the spatial phase inhomogeneity of this mode.

Hydrodynamic considerations of the mode structure of ion-acoustic oscillations in a discharge are given in [7, 9, 10]. However, in contrast to [9, 10], the model of [7] explains the existence of the lowest boundary frequency and demonstrates a good agreement between the calculated and measured frequency of the lowest mode. According to [7], the oscillation modes in the polar coordinates (r, φ) have a structure of the form

$$\delta n_i(r, \varphi) \sim J_m\left(j_{mn}\frac{2r}{d}\right)\cos(m\varphi),$$

where j_{mn} is the nth zero of the mth order Bessel function J_m.

Hence, each mode is characterized by a pair of integers (m, n). The shape of the reconstructed two-dimensional distribution of the lowest mode with two characteristic maxima (Fig. 5a) and the existence of a phase shift of π rad between them agree well only with the theoretical structure of the $m = 1$, $n = 1$ mode (shown in Fig. 5b). This allows us to identify the lowest mode as the (1, 1) mode, which does not contradict the results of [8–10]. Therefore, using the emission tomography method adapted to studying the spatial characteristics of plasma oscillations, the two-dimensional spatial structure of the lowest mode of the ion-acoustic plasma instability of a high-current argon laser was determined for the first time and the mode under study was identified.

REFERENCES

1. V. I. Donin, *High-Power Ion Gas Lasers* (Nauka, Novosibirsk, 1991).
2. V. I. Donin, Zh. Éksp. Teor. Fiz. **62**, 1648 (1972) [Sov. Phys. JETP **35**, 858 (1972)].
3. C. P. Wang and S. C. Lin, J. Appl. Phys. **43**, 5068 (1972).
4. N. P. Gadetskiĭ, Yu. V. Tkach, A. V. Sidel'nikova, and V. P. Zeĭdlits, Ukr. Fiz. Zh. **19**, 931 (1974).
5. H. R. Lüthi and W. Seelig, J. Appl. Phys. **48**, 4922 (1977).
6. G. N. Alferov, V. I. Donin, G. I. Smirnov, and D. A. Shapiro, Kvantovaya Élektron. (Moscow) **8**, 13 (1981).
7. V. I. Donin, D. A. Shapiro, D. V. Yakovin, and A. S. Yatsenko, Zh. Tekh. Fiz. **58**, 80 (1988) [Sov. Phys. Tech. Phys. **33**, 46 (1988)].
8. V. I. Donin, V. A. Ivanov, and D. V. Yakovin, Zh. Tekh. Fiz. **71** (4), 36 (2001) [Tech. Phys. **46**, 400 (2001)].
9. L. C. Woods, J. Fluid Mech. **23**, 315 (1965).
10. H. N. Ewald, F. W. Crawford, and S. A. Self, Phys. Fluids **12**, 303 (1969).
11. V. V. Pikalov and T. S. Mel'nikova, *Tomography of Plasma* (Nauka, Novosibirsk, 1995).
12. C. E. Webb, J. Appl. Phys. **39**, 5441 (1968).
13. G. N. Minerbo, J. G. Sanderson, D. B. van Hulsteyn, and P. Lee, Appl. Opt. **19**, 1723 (1980).

Translated by E. Satunina

Technical Physics, Vol. 48, No. 2, 2003, pp. 209–214. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 75–80.

SOLIDS

On Magnetic Field Penetration into a Granulated Superconductor

M. V. Belodedov and S. V. Chernykh

Volgograd State University, Volgograd, 400062 Russia

e-mail: mic_bel@tele-kom.ru

Received May 18, 2001; in final form May 30, 2002

Abstract—The frequency dependence of the magnetic response of $YBa_2Cu_3O_{7-x}$ high-temperature superconducting ceramics is studied. The experimental results are obtained for different values of the constant ((0–99) × 10^{-4} T) and variable ((0.18–7.24) × 10^{-4} T) components of the external magnetic field in the frequency range from 60 Hz to 1 MHz. A simple qualitative model of the distributed Josephson medium is offered to account for the results obtained. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It is known that if an electromagnetic wave is incident onto a conductor surface, the field penetrates to a skin-layer depth. The skin effect is inherent not only to metals but to superconductors as well. For example, papers [1, 2] consider normal and anomalous skin effects in superconductors in the framework of a two-fluid model.

The pattern of magnetic field penetration in a granulated superconductor is quite complicated (see, for example, [3]) and, apparently, not fully understood [4, 5].

Penetration of the alternative magnetic field in a cylindrical granulated superconductor was studied in [4] using a Hall effect device. This paper reported that the magnetic flux gradient in samples were independent of frequency over all the range of frequencies used from 5 to 500 Hz.

Paper [5] investigated the frequency dependence of the real component of surface impedance in the superconductor samples using special precision measuring equipment. The sample composition was similar to that used in [4]. There, the real component of surface impedance is also independent of frequency in the range from 10 Hz to 1 kHz for a variable component amplitude from 10^{-5} to 10^{-3} T.

EXPERIMENTAL TECHNIQUE

Samples of $YBa_2Cu_3O_{7-x}$ high-temperature superconducting (HTSC) ceramics manufactured by the method of solid phase synthesis were studied. The samples were of cylindrical shape with a diameter of 9 mm and a height of 15 mm. The experimental setup is sketched in Fig. 1.

The setup consists of a *G* (G3-117) sinusoidal voltage generator, a current source *C* (a U7-5 output amplifier loaded with a constant resistance *R* connected in series to a sensor) governed by a generator, a signal sensor *D*, and a selective nanovoltmeter *V* (type 233). The sensor consists of current and signal coils wound around a cylindrical sample of HTSC ceramics one over another. Coils 15 mm in length are wound with a PELShO-0.05 copper wire and consisted of 95 turns each. The current coil is wound over a signal one. Resistance *R* is connected in series to the current coil ($R \gg 2\pi f L_k$ in the range of frequency *f* from zero to $f = 2 \times 10^6$ Hz, where L_k is the current coil inductance).

The external magnetic field is directed along the sample axis. The constant component of the magnetic field H_0 is determined by an external coil (not shown in Fig. 1) in which the sensor is inserted and the variable component $h_0\cos(\omega t + \psi)$ is produced by the signal coil.

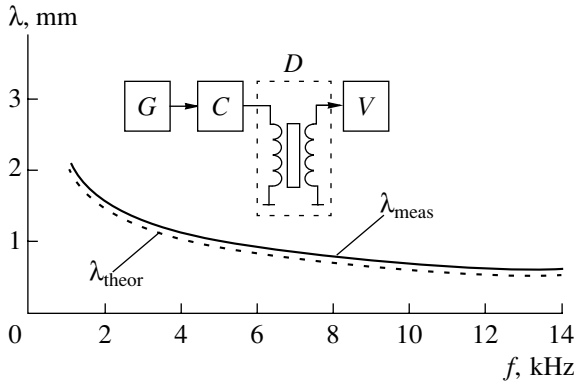


Fig. 1. Schematic view of the experimental setup and reference graphics. λ_{theor} and λ_{meas} are the calculated and measured depth of field penetration into a copper sample.

1063-7842/03/4802-0209$24.00 © 2003 MAIK "Nauka/Interperiodica"

This setup was used to investigate the frequency dependence of the first harmonic of the HTSC sample's response to the external magnetic field $H = H_0 + h_0\cos(\omega t)$ in the frequency range from 60 Hz to 1 MHz for different values of the constant ($H_0 = (0–99) \times 10^{-4}$ T) and variable ($h_0 = (0.18–7.24) \times 10^{-4}$ T) components.

The setup allows one to estimate the penetration depth of the magnetic field into a sample using the formula

$$U_1/U_0 = [\pi r^2 - \pi(r - \lambda_{meas})^2]/\pi r^2 = [2r\lambda_{meas} - \lambda_{meas}^2]/r^2,$$

where U_1 represents the selective voltmeter readings in the presence of a sample; U_0, those in the absence of a sample; and r, the sample radius.

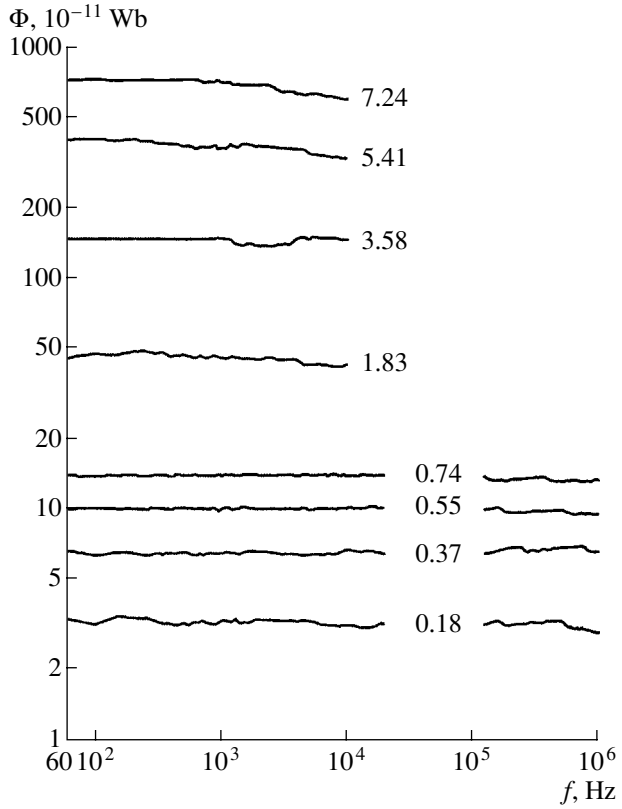


Fig. 2. Frequency dependence of magnetic flux amplitude Φ penetrating into the sample for different values of h_0 at $H_0 = 0$. Figures by curves are values of h_0, 10^{-4} T.

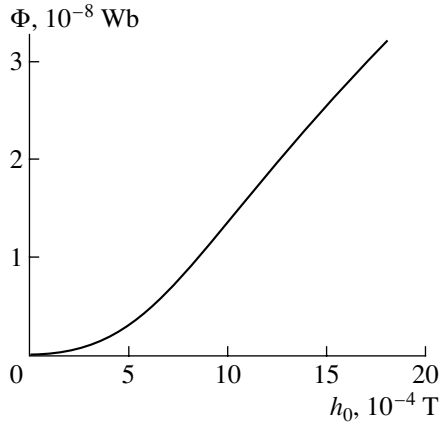


Fig. 3. Dependence of Φ on the variable magnetic field amplitude at $H_0 = 0$ and $f = 2$ kHz.

Calibration tests were carried out for a sample of annealed copper. The frequency dependence of the field penetration depth λ into a copper sample is shown in Fig. 1: λ_{theor} is calculated with reference data [6] and [7], and λ_{meas} is measured using the setup and then calculated by the formula given above.

Estimation of a random measuring error leads to the conclusion that the penetration depth of the magnetic field is measured by the above-described method within an accuracy of 0.03 mm, which corresponds to a relative error of about 3%. In this case, the relative measuring error of the magnetic flux is equal to

$$\frac{\Delta\Phi}{\Phi} = \frac{2\pi B(r-\lambda)\Delta\lambda}{\pi B(2r\lambda - \lambda^2)} = \frac{2r - 2\lambda}{2r - \lambda}\frac{\Delta\lambda}{\lambda}.$$

From this it follows that this value need not be higher than that of the penetration depth of the magnetic field into the sample.

In summary, the above technique for measuring the magnetic flux penetrating into the sample provides a relative measuring error of 3%.

EXPERIMENTAL RESULTS

The amplitude of the magnetic field Φ penetrating into the sample can be derived from the selective voltmeter readings U_1^* as $\Phi = U_1^*/2\pi f N$.

Figure 2 shows the frequency dependence of Φ in different frequency ranges at $H_0 = 0$ for various h_0. In the frequency range from 120 kHz to 1 MHz, the following devices were used: a G4-164 generator, a U7-5 power amplifier, and a V6-10 selective microvoltmeter.

Figure 3 shows the dependence of Φ as a function of the variable magnetic field amplitude h_0 for $H_0 = 0$ and $f = 2$ kHz.

Figure 4 shows the frequency dependence of Φ at the fixed amplitude of the variable field $h_0 = 1.83 \times 10^{-4}$ T in the constant magnetic field $H_0 \approx 0.2 \times 10^{-4}$ and 33×10^{-4} T.

DISCUSSION

In [4], the fact that the magnetic flux gradient in the samples is independent of frequency in the range from 5 to 500 Hz is explained by the absence of viscous forces. According to estimates in [8], the effects of the viscous flow of the vortex structure in HTSC ceramics

become important for frequencies of $\omega_\eta \geq 10^8$ rad/s. The experiments in [4] used frequencies that were many orders of magnitude lower; hence, the viscous forces are reasonably neglected in [4].

In [5], the results obtained are accounted for in the framework of a pinning model and a viscous flow of hypervortices, but, unfortunately, these results have not been confirmed by any estimate.

In this paper, to account for the results obtained and to estimate the magnetic parameters of high-temperature superconductors, an effort is made to draw an analogy between a granulated superconductor and the well-known [9] distributed Josephson junction. Such a model has no pretensions to completeness or mathematical rigorousness, but allows us to qualitatively explain the magnetic processes in HTSC ceramics.

Consider the magnetic field penetration into a cylindrical superconductor of HTSC ceramics. We will consider the ceramics as a set of weakly bound superconducting granules. Granule material is generally recognized to have properties of the type II superconductor with magnetic field penetrating in the form of Abrikosov vortices. A mean granule size L^* is of the same order of magnitude as typical crystallite dimensions of about 10 μm [4]. Hence, the condition $L^* \gg \lambda_L$ is met, where λ_L is the London penetration depth for the granules; its value is about 10^{-5} cm [4] at $T = 77$ K. Let us assume that a number of voids and regions with small crystallites $L^* < \lambda_L$ between large granules can be neglected and extensive (not point) weak bonds (Josephson contacts) prevail. Such a structure is schematically shown in Fig. 5. The first critical field of a granule H_{c1}^g is about $(50–100) \times 10^{-4}$ T at 77 K [10].

Thus, the magnetic field lower than H_{c1}^g penetrates from the surface of a cylinder to its center along the Josephson junctions. Some of these junctions are shown in Fig. 5 by bold solid lines. In this case, one can obviously assume that the experimentally observed effects are related to the magnetic field penetration into the Josephson junction. Let us consider the distributed Josephson junction $-L/2 \leq z \leq L/2$ placed in the external magnetic field (Fig. 6). The uniform magnetic field is assumed to be directed along axis *Y*. As has already been mentioned, the thickness of superconducting electrodes is much greater than the London penetration depth.

Electrodynamics of the extensive Josephson junction can be described by the partial differential equation of a sin-Gordon type [9, 11]. Disregarding dissipation in our one-dimensional case, this equation has the form

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{v_\phi^2}\frac{\partial^2}{\partial t^2}\right)\varphi = \frac{1}{\lambda_j^2}\sin\varphi, \qquad (1)$$

where φ is the phase difference of the order parameter at the Josephson junction; v_φ, the wave phase velocity (Swihart velocity); and λ_j, the Josephson penetration depth.

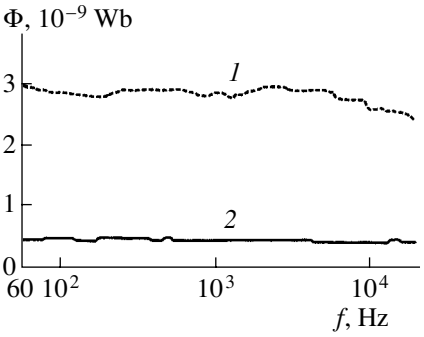


Fig. 4. Frequency dependence of Φ for different values of the constant magnetic field H_0 [(*1*) 33×10^{-4} and (*2*) $\approx 0.2 \times 10^{-4}$ T].

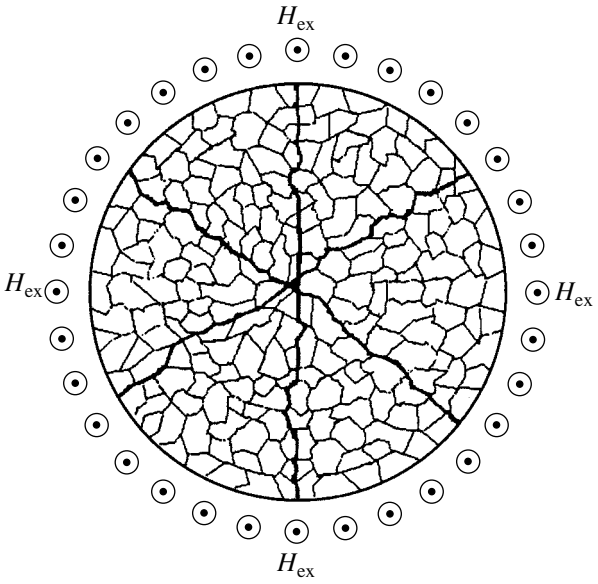


Fig. 5. HTSC ceramics structure in the external magnetic field (draft). Solid lines are extensive weak bonds between granules.

The experimentally observed values are related to the phase differences $\varphi(z, t)$ by the relations [9]

$$H(z) = \frac{\Phi_0}{2\pi d}\frac{d\varphi(z)}{dz}, \quad J(z) = J_c\sin(\varphi). \qquad (2)$$

Here, $H(z)$ is the magnetic intensity in the Josephson junction at a point z in the direction of the Y axis; $d = 2\lambda_L + s$; s is the dielectric thickness (oxide layer); λ_L is the London penetration depth; and $J(z)$ is the superconducting current density in the direction of the X axis.

Let the external magnetic field H_{ex} change by the law $H_{ex} = h_0\cos\omega t$ (here, the H_{ex} and j directions, shown in Fig. 6, can be reversed). Let us determine a stationary

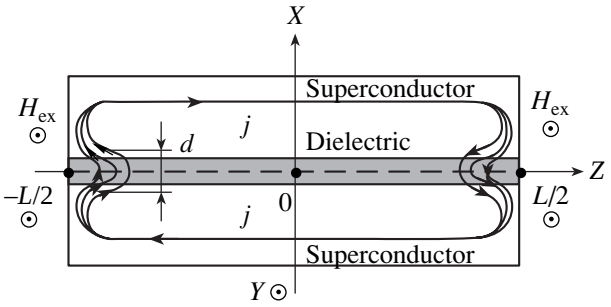


Fig. 6. The distributed Josephson junction in the external magnetic field H_{ex}.

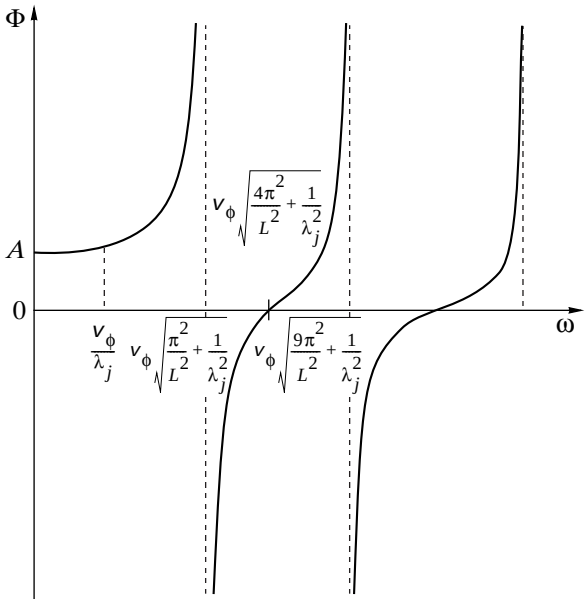


Fig. 7. Frequency dependence of the amplitude of the magnetic flux captured by the distributed Josephson junction, where $A = \Phi(\omega = 0) = 2d\lambda_j H_{ex}\sinh(L/2\lambda_j)/\cosh(L/2\lambda_j)$.

distribution of the magnetic field amplitude $H(z)$ in the junction.

For weak magnetic fields, i.e., for $h_0 \ll \Phi_0/(2\pi\lambda_j d)$ [1], the currents flowing through the junction are small ($J(z) \ll J_c$) and the phase difference φ is small. Hence, Eq. (1) can be expressed as

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{v_\phi^2}\frac{\partial^2}{\partial t^2}\right)\varphi = \frac{\varphi}{\lambda_j^2}. \quad (3)$$

We will express the stationary oscillation regime in the junction in the form

$$\varphi(z, t) = \varphi(z)\cos(\omega t + \psi). \quad (4)$$

Substituting Eq. (4) into Eq. (3) and differentiating, we find

$$\frac{d^2\varphi(z)}{dz^2} = \left(\frac{1}{\lambda_j^2} - \frac{\omega^2}{v_\phi^2}\right)\varphi(z). \quad (5)$$

Let us look for solutions of Eq. (5) with boundary conditions

$$\frac{\Phi_0}{2\pi d}\frac{d\varphi(z)}{dz}\bigg|_{z=-L/2} = \frac{\Phi_0}{2\pi d}\frac{d\varphi(z)}{dz}\bigg|_{z=L/2} = H_{ex}. \quad (6)$$

In the case $\omega^2/v_\phi^2 < 1/\lambda_j^2$ in view of Eqs. (2) and (6), the distribution of the magnetic field amplitude in the junction is equal to

$$H(z) = \frac{\cosh(\xi z)}{\cosh(\xi L/2)}H_{ex}, \quad (7)$$

where $\xi^2 = 1/\lambda_j^2 - \omega^2/v_\phi^2$.

Hence, the amplitude of the total magnetic flux in the junction is

$$\Phi = d\int_{-L/2}^{L/2} H(z)dz = \frac{2d}{\xi}\frac{\sinh(\xi L/2)}{\cosh(\xi L/2)}H_{ex}. \quad (8)$$

In the case $\omega^2/v_\phi^2 > 1/\lambda_j^2$, the distribution of the magnetic field amplitude equals

$$H(z) = \frac{\cos(k_v z)}{\cos(k_v L/2)}H_{ex}, \quad (9)$$

where $k_v^2 = \omega^2/v_\phi^2 - 1/\lambda_j^2$ and the magnetic flux amplitude is equal to

$$\Phi = d\int_{-L/2}^{L/2} H(z)dz = \frac{2d}{k_v}\frac{\sin(k_v L/2)}{\cos(k_v L/2)}H_{ex}. \quad (10)$$

The dependence $\Phi(\omega)$ is shown in Fig. 7. Let us estimate the threshold value of ω providing Eq. (5) with a solution in the form of standing waves. Equating $(k_v)^2 = 0$ and substituting in this relationship the values $v_\phi = 0.05c$ and $\lambda_j \sim 100$ μm [9], where c is the velocity of light in free space, we find $f \approx 2.4 \times 10^{10}$ Hz such that it coincides with the values of the typical plasma frequency given in [9].

As is easily seen from Fig. 2, the dependence of the amplitude of magnetic flux captured by the sample as a function of the exciting field frequency is absent for a field amplitude lower than 2×10^{-4} T. This allows us to assert that the value H_{c1} of HTSC ceramics cannot be higher than 2×10^{-4} T.

Use of a constitutive equation of the granulated superconductor offers a more correct description of its electrodynamics. Such an equation was offered in [12]:

$$\mathbf{j} = -k\mathbf{A}\exp\left(-\frac{|\mathbf{A}|^2}{A_0^2}\right) - a\dot{\mathbf{A}} - b\ddot{\mathbf{A}}, \qquad (11)$$

where **j** is the current density in a superconductor; **A** is the vector potential of the magnetic field **H** = curl**A**; and the positive constants k, A_0, a, and b are determined by the parameters of the granulated superconducting medium.

In weak magnetic fields $|\mathbf{A}| \ll A_0$, Eq. (11) is given by

$$\mathbf{j} = -k\mathbf{A} - a\dot{\mathbf{A}} - b\ddot{\mathbf{A}}.$$

Taking into account the Maxwell equation curl**H** = curlcurl**A** = $(4\pi/c)$**j**, we reduce the given problem of magnetic field penetration into a granulated superconducting medium to the differential equation

$$\frac{\partial^2 A}{\partial z^2} = k^*A + a^*\dot{A} + b^*\ddot{A}.$$

Disregarding the dissipative term $a^*\dot{A}$, this equation is equivalent to Eq. (3) describing magnetic field penetration into the distributed Josephson junction.

The magnetic field penetrates into the granulated superconducting medium into both granules comprising the medium and intergranule spaces [3], which are the distributed Josephson junctions. The field penetrates into intergranule spaces according to the law $H_1 = H_{ex}\exp(-x/\lambda_j)$ [1] in the case of a weak external field $H_{ex} \ll H_{c1}$. Here, H_{c1} is the value of an external field H_{ex}, during the existence of which a vortex at the junction is justified energetically. The magnetic field distribution in granules is described by the expression $H_2 = H_{ex}\exp(-x/\lambda_L)$ for a weak external magnetic field $H_{ex} \ll H_{c1}^g$. The magnetic field induction in a granulated medium (i.e., the field H averaged over volume involving a large number of granules) is determined as

$$B = \alpha H_1 + \beta H_2$$
$$= H_{ex}(\alpha\exp(-x/\lambda_j) + \beta\exp(-x/\lambda_L)),$$

where α and β are the contributions in the total volume from intergranule spaces and granules, respectively.

Since $\alpha < \beta$, the magnetic field induction should have the distribution shown in Fig. 8. As is seen, the magnetic field jumps at the boundary of a granulated superconducting medium. This magnetic field jump was experimentally observed, for example, in [4]. It is of interest that in [4] the magnetic field was measured by a Hall effect device with working-area sizes of 50×50 μm and a thickness of 180 μm, which is much larger than the granule size (~10 μm [4]). These experimental results can be considered as an additional indirect corroboration of the model we offer.

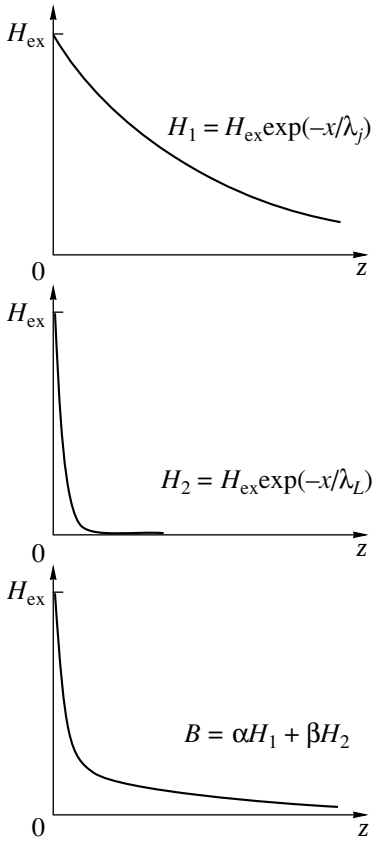


Fig. 8. Magnetic field penetration into the granulated superconducting medium.

CONCLUSION

The considered analogy between a Josephson medium and distributed Josephson junction allows us to explain a number of properties of granulated superconductors, the most important of which is the frequency independence of magnetic response of the HTSC sample found earlier. The obtained estimate of an upper boundary frequency of electromagnetic field penetration into an HTSC (about 10 GHz) should be taken into account in designing, for example, a compensation HTSC magnetometer [13], in which an increase in the operating frequency should increase its sensitivity.

REFERENCES

1. V. V. Shmidt, *Introduction to the Physics of Superconductors* (Nauka, Moscow, 1982).
2. D. Mattis and J. Bardeen, Phys. Rev. **111**, 412 (1958); D. Mattis and J. Bardeen, in *Theory of Superconductivity*, Ed. by N. N. Bogolyubov (Inostrannaya Literatura, Moscow, 1960), p. 172.

3. É. B. Sonin, Pis'ma Zh. Éksp. Teor. Fiz. **47** (8), 415 (1988) [JETP Lett. **47**, 496 (1988)].
4. V. M. Dzugutov, N. A. Podlevskikh, and L. M. Fisher, Sverkhprovodimost: Fiz., Khim., Tekh. **3** (1), 52 (1990).
5. V. K. Ignat'ev and I. V. Neginskiĭ, Fiz. Nizk. Temp. **26** (4), 340 (2000) [Low Temp. Phys. **26**, 247 (2000)].
6. V. V. Nikol'skiĭ and T. I. Nikol'skaya, *Electrodynamics and Propagation of Radiowaves* (Nauka, Moscow, 1989).
7. *Handbook on Materials for Electrical Engineering*, Ed. by Yu. V. Koritskiĭ, V. V. Pasynkov, and B. M. Tareev, 3rd ed. (Énergoatomizdat, Leningrad, 1988), Vol. 3.
8. É. B. Sonin and A. K. Tagantsev, Zh. Éksp. Teor. Fiz. **95** (3), 994 (1989) [Sov. Phys. JETP **68**, 572 (1989)].
9. A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982; Mir, Moscow, 1984).
10. *Physical Properties of High Temperature Superconductors*, Ed. by D. M. Ginsberg (World Sci., Singapore, 1989; Mir, Moscow, 1990).
11. T. van Duser and C. W. Turner, *Principles of Superconductive Devices and Circuits* (Elsevier, New York, 1981; Radio i Svyaz', Moscow, 1984).
12. M. V. Belodedov and V. K. Ignat'ev, Sverkhprovodimost: Fiz., Khim., Tekh. **4**, 1661 (1991).
13. V. K. Ignat'ev and S. V. Chernykh, Prib. Tekh. Eksp., No. 2, 124 (1996).

Translated by M. Astrov

Technical Physics, Vol. 48, No. 2, 2003, pp. 215–218. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 81–86.

SOLIDS

Fractal Characteristics of the Surface of Tungsten Carbide–Manganese Steel Composite after Plastic Deformation

I. N. Sevostjanova and S. N. Kulkov

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

e-mail: kulkov@ms.tsc.ru

Received May 22, 2002

Abstract—The structure evolution in WC–80G4 steel composite heavily loaded by compression is studied. The fractal properties of the deformation relief appearing on the surface are determined. The fractal dimension of the deformed surface profile is shown to depend on the mean spacing between bands of localized deformation and on the parameters of the material fine crystal structure. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Deformation usually causes a surface relief to appear in both ductile and brittle metals and alloys such as hard alloys. However, relationships between the relief and structural properties of brittle materials are little understood. A promising approach to studying the relief on the surface of the deformed material is to analyze the profile of the signal of secondary electrons escaping the surface under investigation in a scanning electron microscope [1–4], since the number of secondary electrons is uniquely related to the surface relief. Moreover, the deformation relief on the material surface is the result of structural changes occurring at various scale levels. In particular, when studying WC–80G4 steel composite, we showed [4] that the fractal dimension (*Df*) measured at various scale levels linearly depends on the microstructure parameters of a binder, namely, coherently diffracting domains (CDDs) and microstrained lattice regions. However, the relation between the parameters of the fine crystal structure of phases making up the composite and the fractal characteristics of its surface has not been studied.

The purpose of this work is to determine the fractal dimension of the surface and trace its evolution upon loading, as well as to establish a relation between the fractal dimension and micro- and macroparameters of the structure of the composite.

EXPERIMENTAL

We studied WC–80G4 steel composite material with an 80% strengthening phase. The production of the hard alloy was described elsewhere [5]. The mean size of a carbide grain in the composite was 2.4 ± 0.2 μm, the mean size of intercarbide layers was 1.2 ± 0.2 μm, and the general porosity was smaller than 0.5%. Test specimens measured $0.004 \times 0.005 \times 0.007$ m and had a polished side face. An INSTRON-1185 machine was used to load specimens in 100-MPa steps up to fracture. After each loading step, the sample was unloaded and its structure and surface relief were studied at various scale levels. Metallographic examination was performed with a NEOPHOT-21 optical microscope using direct and oblique illumination. The mean spacing between bands of localized deformation was determined by the linear intercept method [6]. X-ray diffraction analysis was carried out with a DRON-UM1 X-ray diffractometer using filtered CuK_α radiation. The size of CDDs was calculated using the (111) reflection from the γ phase and (100) reflection from WC; the lattice microstrains, with the (222) reflection from the γ phase and (300) from WC. We chose these reflections, assuming that the major contribution to the broadening of lower index diffraction lines is from CDDs and of higher index lines, from lattice microstrains [7]. The surface relief was analyzed by the determination of the fractal dimension (*Df*) using a RÉM-200 scanning electron microscope at an accelerating voltage of 30 kV by the technique described in [2].

RESULTS AND DISCUSSION

The metallographic examination of the evolution of the WC–80G4 steel composite surface upon compression loading showed no visible changes at stresses of 0–600 MPa. At stresses from 700 to 1700 MPa, two processes proceed simultaneously: the deformation of large binder regions, showing up as the relief between carbide grains, and the appearance of slip bands in a number of coarse carbide grains. At this loading stage, carbide grains and the binding phase are deformed simultaneously without breaking the continuity of the material. At stresses of 1700–2200 MPa, the number of

1063-7842/03/4802-0215$24.00

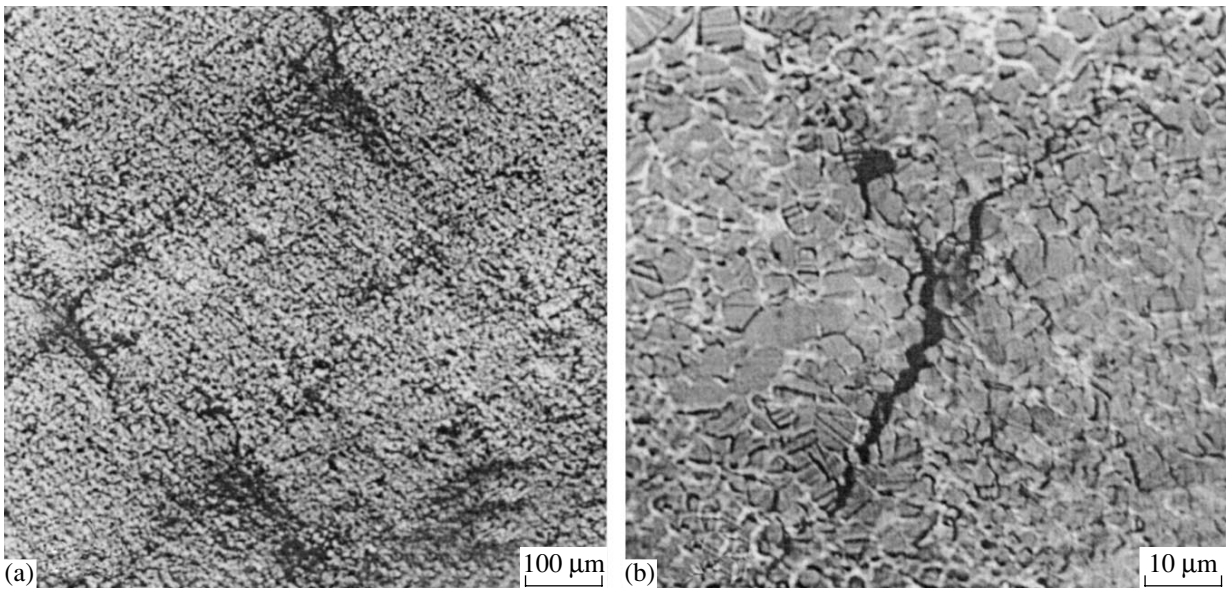


Fig. 1. Micrographs of (a) bands of localized deformation on the surface of the hard alloy and (b) macrocracks originating at the sites of band intersection.

plastically deformed grains on the composite surface increases and finer grains are involved in the plastic flow. The binder in narrow intercarbide spacings undergoes severe deformation. Along with the deformation of the structure constituents, individual cracks of size comparable to the carbide grain size appear. These are placed predominantly along the carbide–binder interface and randomly distributed over the composite surface. An increase in the stress to 3000 MPa leads to an increase in the density of cracks and the amount of carbide grains deformed.

At 3000–3900 MPa, bands of local plastic deformation appear on the composite surface (Fig. 1a). These bands are regions with a large number of deformed carbide grains (with slip traces) and the severely deformed binding phase. Moreover, cracks running along the carbide–carbide and carbide–binder interfaces are found to coalesce in bands of localized deformation. Under severe deformation, individual carbide grains, as a rule coarse ones, break up into fragments. The bands are oriented in two directions and extend in the direction of maximum shear stresses at an angle of about 45° to the loading axis, irrespective of the structure of the material. Their length varies in a wide range from 20 μm to 4 mm, and their width is 5–10 μm, i.e., two to four times as large as the carbide grain size.

First, the bands of local deformation spaced at 300–700 μm appear. Their length is comparable to the specimen dimensions, and their width increases with plastic strain. Shorter bands of localized deformation form between these large bands. Their intersection leads to the fragmentation of the entire specimen surface. Before the composite fails, the stresses accumulate basically in several adjacent bands whose length is comparable to the specimen dimensions. As they intersect, macrocracks 40–150 μm in size that are oriented along the bands of localized deformation form (Fig. 1b). These macrocracks are the precursors of hard alloy fracture.

Figure 2 (curve *1*) shows the variation of the mean interband spacing with residual strain. The metallographic examination of the surface demonstrates that the intense formation of bands of localized deformation, which shrinks the interband spacing, takes place up to a residual strain of 3.5%. At final stages of loading, the deformation of the structure constituents, prefracture processes in the form of crack initiation and coalescence along the interfaces, and carbide fragmentation occur in several adjacent bands that have the most pronounced surface relief. This means that new bands do not arise and that a catastrophic crack, which causes composite fracture, extends via the coalescence of several macrocracks in adjacent bands of localized deformation.

The stress–strain curve for the alloy obtained by measuring the residual strain after unloading the specimen is typical. The strain increases linearly up to σ = 2500 MPa. From 2500 MPa to fracture, the strain is a parabolic function of the stress. Plotted in the logarithmic coordinates, the σ–ε curve exhibits three linear segments. They correspond to three stages of the strain hardening of the composite material, which reflect radical modifications of the composite structure and, possibly, changes in deformation mechanisms. In the first segment with a strain-hardening exponent $n = 0.54$, the matrix and the carbide phase are deformed simultaneously without visible signs of fracture. The second segment ($n = 0.31$) (macroyield) corresponds to the

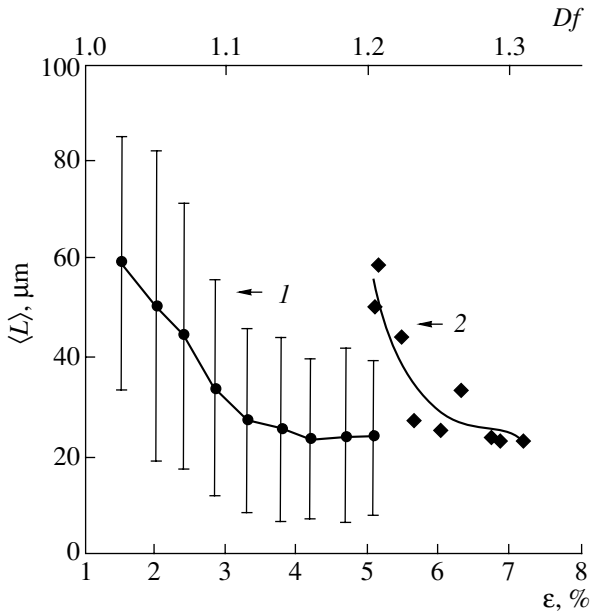


Fig. 2. Mean spacing between the bands of localized deformation as a function of the (*1*) residual strain and (*2*) fractal dimension.

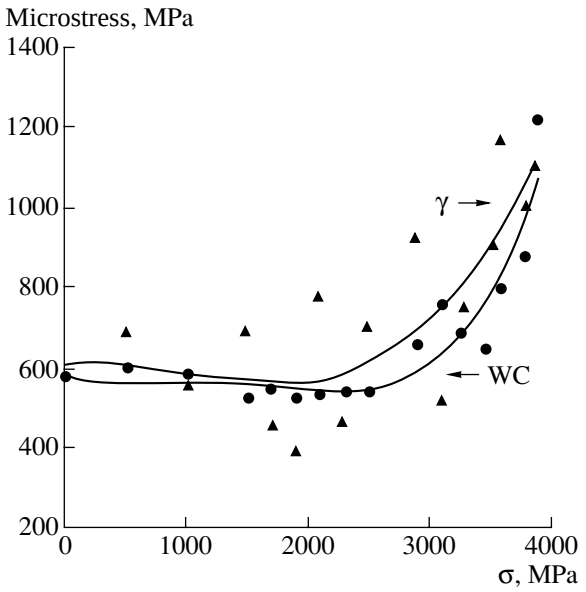


Fig. 3. Microstresses in the binding phase and tungsten carbide vs. the applied stress.

intermediate region. In the third segment (n = 0.15), processes leading to the fracture of the alloy (the formation and development of bands of localized deformation and micro- and macrocracks) accompany the deformation of the structure constituents.

Using the X-ray diffraction data, we measured the sizes of CDDs and microstrains $\langle\varepsilon^2\rangle^{1/2}$ in the matrix and carbide lattice. The initial γ phase has a fine-grained structure with a block size of ≈60 ± 5 nm. The CDD size remained unchanged up to stresses of 600 MPa. As the applied stress grows (σ = 700–1700 MPa), the CDD size in the austenite rises from 60 to 67 nm. This increase in the block size in the austenitic matrix at the early stage of loading may be due to a decrease in the density of dislocations because of their annihilation and arrival at grain boundaries [8]. At stresses above 2000 MPa, the CDD size decreases to 30 nm and $\langle\varepsilon^2\rangle^{1/2}$ increases.

The size of CDDs in the tungsten carbide is initially 400 ± 50 nm. Up to stresses of ≈2500 MPa, the CDD size and $\langle\varepsilon^2\rangle^{1/2}$ in the tungsten carbide remain unchanged. At higher stresses, the size of CDDs decreases to 100 nm.

It was found that, although the lattice microstrains are different in the austenitic matrix and tungsten carbide, the generating microstresses are equal at all stages of loading the composite material (Fig. 3). This indicates the absence of a stress jump at the carbide–binder interface and means that the binder and strengthener are plastically deformed as a whole and interfacial cracks are due to the limited plastic deformation of the tungsten carbide.

Figure 4 illustrates the fractal dimension of the deformed surface profile during the steplike loading of the specimen as a function of the residual strain and applied stress. As the strain increases, *Df* varies linearly. At the same time, the fractal dimension varies nonmonotonically with stress: increasing and decreasing segments alternate.

The three stress ranges in the σ–ε curve, which reflect radical changes in the composite structure during deformation were compared with the fractal dimensions. These ranges are separated by vertical lines in Fig. 4. In the first range, the fractal dimension first increases and then returns to its initial value. The slight difference in *Df* here may be associated with the appearance of deformed regions in the binder and slip bands in coarse carbide grains, as well as with an increase in the CDD size in the austenitic matrix. The increase in the fractal dimension in the second range can be caused by several factors. At this stage, carbide grains and the binding phase undergo severe deformation and finer carbide grains are involved in the process of plastic flow. The number of cracks at the carbide–binder interface increases. The examination of the composite fine structure showed that the block size in the matrix and carbide phase decrease slightly at this stage. In the third range, the fractal dimension first remains unchanged and then, above 3500 MPa, abruptly increases. Structure modifications that occur at final deformation stages imply that plastic deformation competes with composite fracture. On the one hand, many micro- and macrocracks appear, carbide grains become fragmented, and strains localize in several adjacent bands. On the other hand, the sizes of CDDs sharply

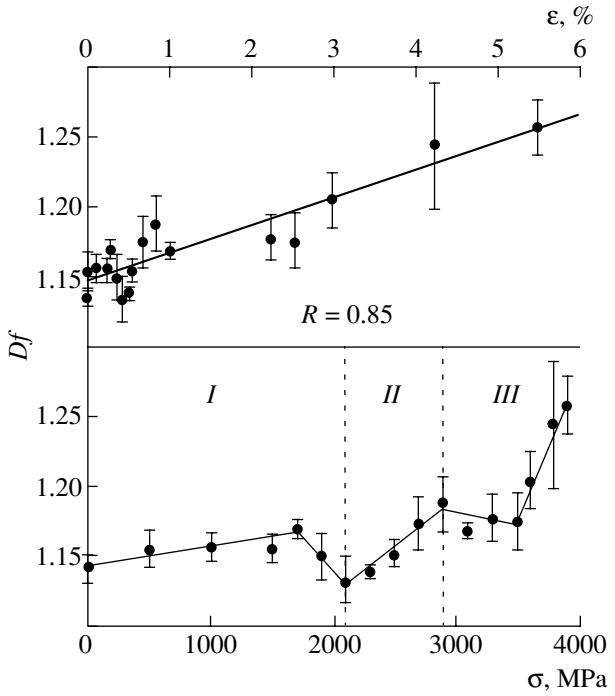


Fig. 4. Variation of the fractal dimension of the surface of the WC–80G4 steel hard alloy with stress and residual strain.

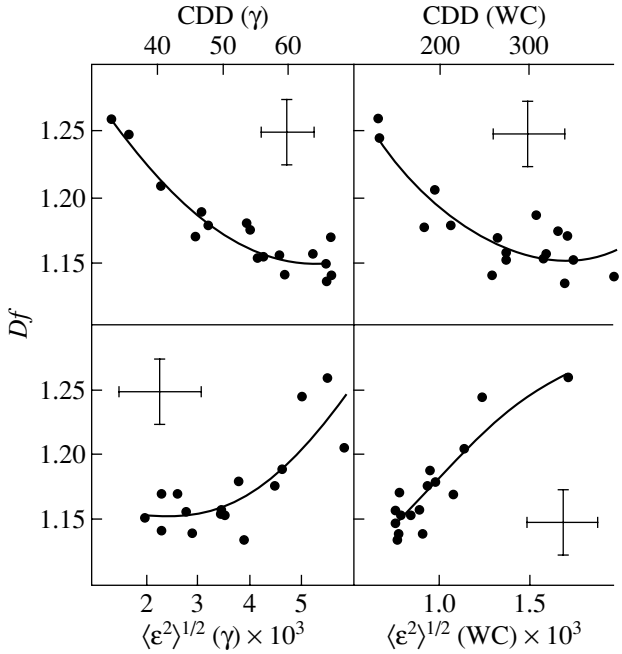


Fig. 5. Relief parameters vs. parameters of the hard alloy fine structure.

decrease and the lattice microstrains in the matrix and carbide phase sharply increase, indicating severe plastic deformation in the structural constituents of the alloy.

Figure 2 (curve *2*) shows the variation of the fractal dimension with the mean spacing between the bands of localized deformation. A decrease in the mean spacing is seen to significantly increase the fractal dimension. The value of *Df* is small where the interband spacing is maximum; that is, the stronger the surface fragmentation, the higher the fractal dimension.

Figure 5 illustrates the dependence of the fractal dimension on the parameters of the fine structures of the binder and carbide phase. The fractal dimension grows with decreasing CDD size in the austenitic matrix and carbide phase. At the same time, an increase in the lattice microstrains in both phases also raises *Df*; in other words, changes in the fractal dimension reflect those in the microstructure parameters.

CONCLUSIONS

(1) A change in the fractal dimension is caused by structural features appearing on the deformed surface, such as slip bands in carbide grains, the relief on the binder, micro- and macrocracks, and bands of localized deformation.

(2) The fractal dimension increases linearly with the residual strain, whereas its stress dependence is nonmonotonic. Three ranges of *Df* variation, which characterize the sequential stages of composite deformation, are revealed.

(3) The deformation relief on the alloy surface is associated with microstructural changes in the material. The fractal dimension grows as the size of coherently diffracting domains in the austenitic matrix and tungsten carbide decreases and lattice microstrains in them increase.

REFERENCES

1. V. S. Ivanova, A. S. Balankin, I. Zh. Bunin, and A. A. Oksogoev, *Synergism and Fractals in Materials Science* (Nauka, Moscow, 1994).
2. P. V. Korolev and S. N. Kul'kov, Perspekt. Mater., No. 3, 21 (1997).
3. V. E. Panin, P. V. Kuznetsov, E. E. Deryugin, *et al.*, Fiz. Met. Metalloved. **84** (2), 118 (1997).
4. I. N. Sevostjanova and S. N. Kulkov, Pis'ma Zh. Tekh. Fiz. **25** (2), 34 (1999) [Tech. Phys. Lett. **25**, 56 (1999)].
5. I. N. Sevostjanova, S. F. Gnyusov, A. P. Garms, and S. N. Kulkov, Perspekt. Mater., No. 4, 37 (1998).
6. S. A. Saltykov, *Stereometric Metallography* (Metallurgiya, Moscow, 1970).
7. Ya. D. Vishnyakov, *Advanced Methods in Studying the Structure of Deformed Crystals* (Metallurgiya, Moscow, 1975).
8. J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1967; Atomizdat, Moscow, 1972).

Translated by K. Shakhlevich

Technical Physics, Vol. 48, No. 2, 2003, pp. 219–224. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 87–92.

SOLID-STATE ELECTRONICS

Effect of Hydrogen on the Photoelectric Properties of Palladium/Anodic Oxide/Gallium Arsenide Schottky Diodes

S. V. Tikhov, E. L. Shobolov, V. V. Podol'skiy, and S. B. Levichev

Lobachevsky State University, pr. Gagarina 23, Nizhni Novgorod, 603950 Russia

e-mail: fdp@phys.unn.ru

Received April 24, 2002

Abstract—The effect of hydrogen on the photovoltage and *I–V* characteristics of palladium/anodic oxide/gallium arsenide Schottky diodes is studied. The oxide thickness that is optimal in terms of the hydrogen sensitivity of the diodes and the depth of atomic hydrogen penetration into the oxide are determined. The mechanism behind the hydrogen effect consists in the chemical adsorption of atomic hydrogen on the gallium arsenide surface, which decreases the barrier height and increases the recombination component of the current. It is shown that a thin tunnel-transparent anodic oxide film is nonuniform in thickness and that hydrogen exposure raises the probability of tunneling through the oxide. It is found that the method of hydrogen detection from the photovoltage response offers a higher sensitivity and detectivity but has a lower speed than the reverse current method. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Schottky diodes (SDs) with a thin tunnel-transparent insulating layer between a catalytic rectifying contact (Pd) and semiconductor are promising for high-speed low-power hydrogen sensors [1]. The effect of hydrogen on the properties of Si diodes with a specially grown or native tunnel-transparent oxide layer is rather well understood [2–6]. GaAs diodes have been less investigated [7–9]. Most of the studies concerned diodes with a native or thermally grown oxide layer of fixed thickness. In this paper, we investigate the effect of hydrogen on the photovoltage and *I–V* characteristics of GaAs diodes with an anodic oxide (AO) layer of various thickness.

EXPERIMENTAL

The diodes were made of *n*-GaAs epitaxial layers with an electron density $n_0 \approx 3 \times 10^{16}$ cm^{-3} grown by atmospheric-pressure metal-organic vapor-phase epitaxy (MOVPE) on GaAs(100) substrates misoriented by 3°. The electron density in the substrate was $n_0 \approx 10^{16}$ cm^{-3}. AO films were grown by anodization in an electrolyte consisting of a 3% aqueous solution of tartaric acid and ethylene glycol (1 : 2). The thickness of the AO films, d_{AO}, was varied from 0 to 20 nm. A semitransparent Pd electrode of area $\approx 8 \times 10^{-3}$ cm^{-2} and thickness ≈10 nm (a transmission of about 80%) was deposited on the unheated oxidized surface of the epitaxial film by vacuum evaporation. The thickness of the AO and Pd films was conventionally determined from the anodization constant and by microweighing, respectively, as well as by a Topometrix® atomic force microscope (AFM) with an accuracy of 0.3 nm using an etched step. The microscope was also used to examine the surface microrelief of the epitaxial and oxide layers in the noncontact mode.

We also studied the open-circuit photovoltage upon illuminating by undecomposed light from a tungsten incandescent lamp, photovoltage spectra with photon energies between 0.6 and 1.5 eV, and *I–V* characteristics of the SDs at temperatures from 300 to 420 K in air and in an air–argon flow with a hydrogen concentration $P_H \approx 0.04$–0.22 vol % [8].

RESULTS AND DISCUSSION

Photovoltage. We found that at temperatures from 360 to 390 K, the hydrogen sensitivity and speed of the diodes were the highest and the change in their properties was reversible.

In an hydrogen-containing atmosphere, the photovoltage V_{ph} always decreased, in agreement with published data, because of the hydrogen-related decrease in the GaAs surface barrier [8, 9].

Figure 1 (curves *1*, *2*) shows the dependences of the photovoltage V_{ph} and its variation ΔV_{ph} due to the hydrogen effect on the oxide layer thickness. It is obvious that both V_{ph} and ΔV_{ph} grow as the AO thickness increases to 3.0 nm. Then, V_{ph} saturates, whereas ΔV_{ph} passes through a maximum (at 3.0–5.0 nm) and drops to zero at $d_{AO} \approx 16$ nm. The increase in V_{ph} (curve *1*) is explained by the formation of the negative charge in the near-surface AO layer during GaAs anodic oxidation [10]. The sensitivity to hydrogen shows up only after the GaAs surface has been covered by a thin oxide layer (curve *2*) and then increases with oxide thickness,

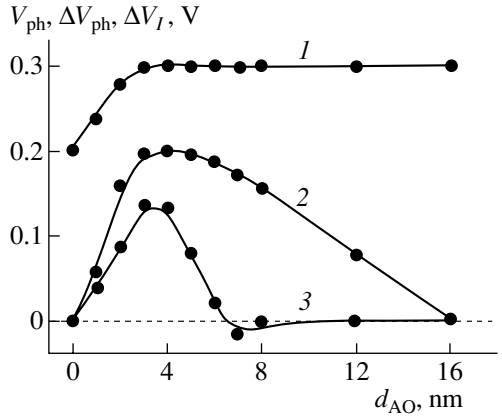


Fig. 1. (*1*) V_{ph}, (*2*) ΔV_{ph}, and (*3*) ΔV_I vs. oxide layer thickness. $T \approx 390$ K, $P_H \approx 0.22\%$.

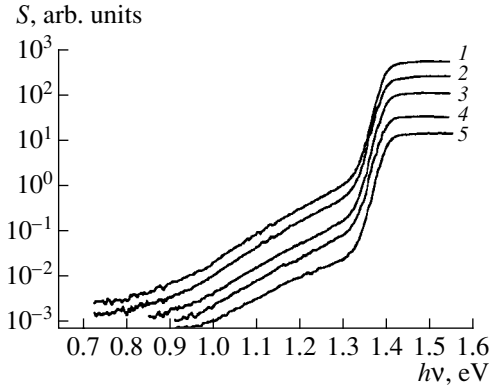


Fig. 2. Photosensitivity vs. photon energy. $d_{AO} \approx 3.4$ nm, $T \approx 366$ K. P_H = (*1*) 0, (*2*) 0.021, (*3*) 0.084, (*4*) 0.17, and (*5*) 0.42%.

because the fraction of the hydrogen-insensitive GaAs–Pd compound diminishes [11].

The decline of ΔV_{ph} as the oxide layer grows while V_{ph} remains constant indicates that the amount of atomic hydrogen reaching the Pd/GaAs interface decreases. Thus, the value $d_{AO} \approx 16$ nm, at which the hydrogen sensitivity virtually disappears, characterizes the depth of hydrogen penetration into the AO at the temperature of measurement (390 K).

A shift ΔV_I of the forward-bias branch of the *I–V* characteristic at a fixed current can serve as another measure of the SD sensitivity to hydrogen [4]. Curve *3* shows that ΔV_I declines at lesser oxide thicknesses than ΔV_{ph} and even changes sign at $d_{AO} \approx 7$ nm, testifying that the forward current decreases under the action of hydrogen.

It was found that, for the light intensity used in the measurements, the photovoltage varies in accordance with the variation of the barrier in the semiconductor. The fact that ΔV_I responds to hydrogen more weakly and drops with d_{AO} more sharply than ΔV_{ph} is naturally explained by the decrease in the current (and hence in its hydrogen-sensitive component) when the oxide resistance grows.

The results which follow refer to Schottky diodes with $d_{AO} \approx 3.4$ nm, which have the maximal sensitivity to hydrogen in terms of both photovoltage and current.

Since the variation of the Schottky contact work function usually changes the GaAs surface condition only slightly [12], the decrease in the Pd/GaAs barrier on exposure to hydrogen can be related either to the decrease in the proton concentration at the AO/GaAs interface due to atomic hydrogen chemisorption or to the decrease in the density of negatively charged surface states (SSs) at this interface due to hydrogen passivation. In order to decide between these mechanisms, it was necessary to find a way of evaluating the density of negatively charged surface states below the equilibrium Fermi level and to make sure that hydrogen has an effect on them. According to [13], surface states below the equilibrium Fermi level show themselves as a characteristic extended portion of extrinsic photosensitivity $S(V_{ph}/L)$, which increases exponentially with the photon energy $h\nu$ and is observed in the gallium arsenide/electrolyte system. We carried out similar measurements with our diodes.

Figure 2 shows typical spectra $S(h\nu)$ measured in air and in air–argon flows with different hydrogen concentrations. The wide extrinsic photosensitivity band is seen in the energy range of E_c – (0.72–1.32 eV), where the value of S increases almost exponentially with $h\nu$ due to the response from the surface states. The small convexity of the exponential is associated with the response of bulk defects in GaAs that are generated by Pd–GaAs chemical interaction [14]. This convexity was not observed in the spectra for diodes with an Au electrode and for those with an electrolyte contact [13].

It is apparent that the hydrogen exposure reduces S by nearly two orders of magnitude. At the same time, the extrinsic (impurity) photoresponse from the surface states decreases insignificantly with respect to the intrinsic photoresponse from GaAs. Hence, the density of surface states lying below the Fermi level, including the density of acceptor surface states, does not respond to hydrogen exposure: the barrier in the semiconductor lowers as a result of hydrogen chemisorption at the AO/GaAs interface with the formation of the proton charge. This conclusion is in agreement with that drawn in [9] for hydrogen-sensitive Schottky diodes with a thermal oxide.

***I–V* characteristics and temperature dependences of the current.** Figure 3 shows the *I–V* characteristics of the diodes taken at ≈390 K in the air flow (curves *1*, *2*) and in the air–argon flow with a different H_2 content (curves *3–5*). Evidently, the hydrogen exposure increases the current density j (by one to two orders of magnitude) and thus considerably modifies

the shape of the *I–V* characteristic. The forward-bias portion of the *I–V* curve deviates markedly from an exponential even at low voltages, tending toward saturation at high forward biases ($V > 0$ at Pd). The reverse-bias portion almost saturates with an increase in negative V unlike the dependence $j \sim \sqrt{V}$, which is observed in the air flow (cf. curves *1* and *4*, *5*).

Usually the analysis of the *I–V* characteristic of Schottky diodes with a thin tunnel-transparent oxide layer includes the thermionic emission over the barrier and tunnel emission through the oxide [9, 15]:

$$j = AT^2 T_n(V) \times \exp\left(-\frac{q\varphi_s(V) + qV + E_c - F}{kT}\right)\left[\exp\left(\frac{qV}{kT}\right) - 1\right], \quad (1)$$

where A is the Richardson–Schottky constant, $T_n(V)$ is the tunneling probability, q is the electron charge, k is the Boltzmann constant, $\varphi_s(V)$ is the surface potential of a semiconductor, and $E_c–F$ is the gap between the conduction band bottom E_c and equilibrium Fermi level F in the bulk of the semiconductor.

To elucidate the current transport mechanism and the effect of hydrogen on the diodes, we took the temperature dependences of the forward, j_f, and reverse, j_r, current densities at various voltages and hydrogen concentrations (Fig. 4). They all obey the exponential law in accordance with (1) and exhibit high activation energies, which vary with the sign and magnitude of the bias and with the hydrogen concentration.

The activation energy E_r of the reverse current was independent of the bias voltage and was found to be about 1.0 eV in the air flow. This value is close to the Pd/GaAs barrier height $q\varphi_b = q\varphi_s + E_c - F \approx 1.1$ eV. The value of φ_s was determined from the voltage dependence of the barrier capacity.

Exposure to hydrogen with a concentration of 0.044% increases the reverse current by more than one order of magnitude, leaving the value of $E_r \approx q\varphi_b$ unchanged. This result can be explained by proton accumulation due to the chemisorption of atomic hydrogen at the oxide–semiconductor interface, which thins the tunneling barrier and thereby increases the probability of tunneling through the insulating layer [15].

A further increase in P_H to 0.22% results in the smooth decrease in E_r to 0.65 eV (cf. curves *3*, *4*, and *5*). The maximal change in the activation energy was found to be $\Delta E_r \approx 0.35$ eV, which is probably equal to the change in the barrier height at the Pd/GaAs interface after hydrogen exposure.

The activation energy E_f of the forward current decreases with increasing bias V and P_H (Fig. 4; curves *1*, *2*). The maximal decrease in ΔE_f on exposure to hydrogen with $P_H \approx 0.22\%$ was ≈0.19 eV.

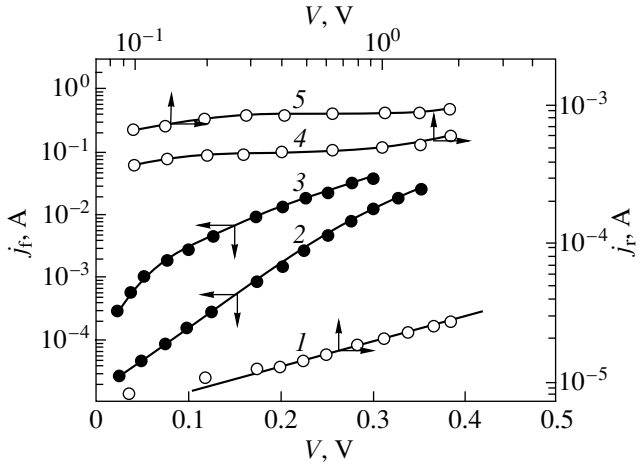


Fig. 3. The effect of hydrogen on the (*2*, *3*) forward-bias and (*1*, *4*, *5*) reverse-bias branches of the *I–V* characteristic: (*1*, *2*) in air flow and (*3–5*) in air–argon flow with H_2. P_H = (*3*, *5*) 0.22 and (*4*) 0.022%; $T \approx 390$ K; $d_{AO} \approx 3.4$ nm.

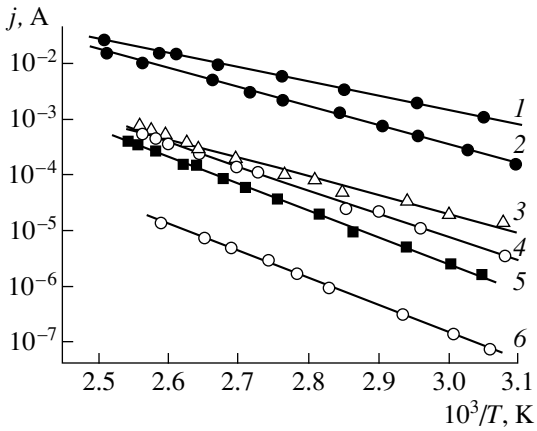


Fig. 4. Temperature dependences of the (*1*, *2*) forward and (*3–6*) reverse currents. V = (*1*, *2*) +0.3 and (*3–6*) –1 V. P_H = (*2*, *6*) 0, (*1*, *3*) 0.22, (*4*) 0.11, and (*5*) 0.044%. $d_{AO} \approx 3.4$ nm.

The general feature of the temperature dependences of j_r and j_f is that the current variation at a fixed temperature is considerably lower than one might expect from the variation of the energy of activation both on exposure to hydrogen and with forward bias on the diode. For instance, as follows from Fig. 4, hydrogen exposure at $T \approx 333$ K increases j_f by a factor of 4, while from the decrease in E_f by 0.19 eV, one could expect an increase in J_f by a factor of 785. For j_r, the increase is 13.6 times versus the expected 2.2×10^5 times. Within the model accepted, it is reasonable to explain this difference by the influence of the tunnel-transparent insulating spacer.

According to (1), the ratio of an experimentally observed current increase to the expected one for a given decrease in the activation energy on exposure to hydrogen at a fixed temperature and voltage is equal to

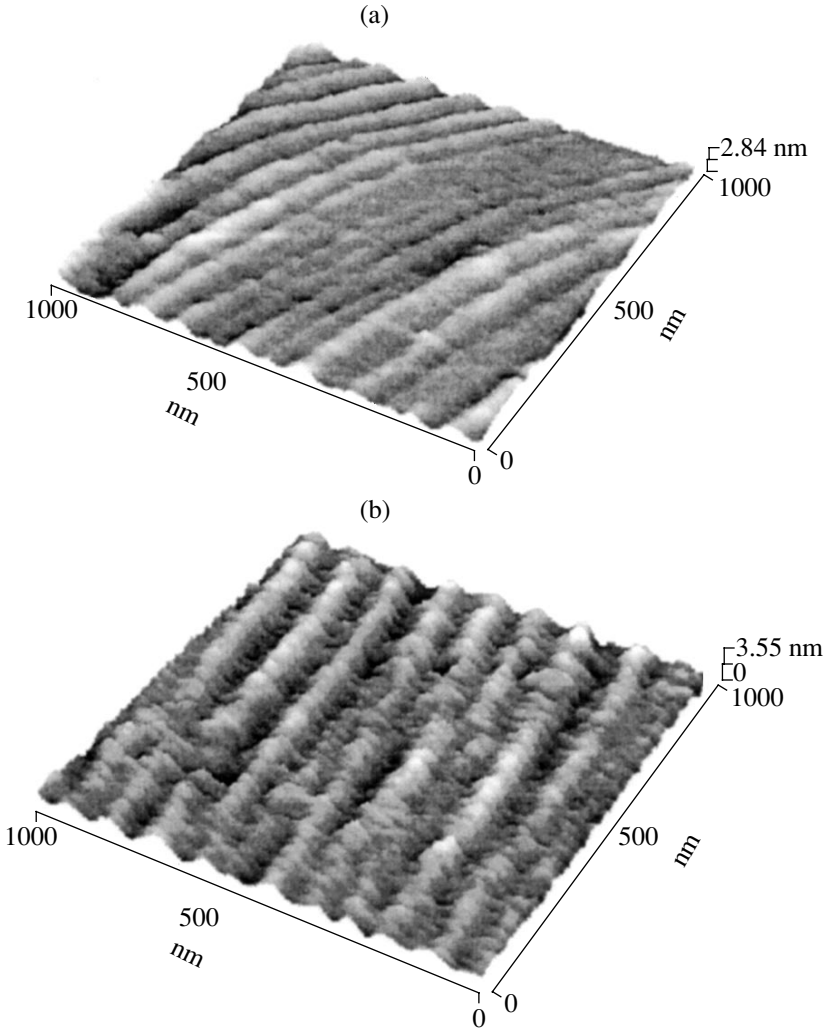


Fig. 5. AFM images of the (a) epitaxial layer and (b) oxide layer of rated thickness 3.4 nm.

the probability of tunneling through the oxide. Then, for a reverse-biased diode at 333 K and $V = -1$ V, $T_n \approx 6.2 \times 10^{-5}$, and calculations by the formula [15]

$$\eta_0 = [\ln(T_n)]^2/(100d_0), \qquad (2)$$

where η_0 is the effective tunnel barrier height in electronvolts and d_0 is the barrier thickness in nanometers, yield $\eta_0 \approx 0.08$ eV if the effective mass of an electron is taken to be equal to the rest mass of an electron in a vacuum and $d_0 = d_{AO}$.

The same value of η_0 is obtained for the forward-biased diode. It differs dramatically from the AO/GaAs barrier height calculated in [16]. Using (2), we have $d_0 \approx 0.6$ nm at $\eta_0 = 1.6$ eV. It is known that anodic oxide films of small thickness have areas of still smaller thickness, where pores are filled [17]. The effective oxide thickness is actually smaller than the nominal thickness d_{AO} because of these thinnings, which may introduce an error in the tunneling barrier height calculated by (2).

The thinnings in the oxide can be observed and measured with an atomic force microscope by comparing the microreliefs of the initial and oxidized surfaces of the GaAs epitaxial layer. The AFM images of these surfaces are presented in Fig. 5. On the initial surface (Fig. 5a), growth steps with incipient hillocks along the length are seen, whereas the surface covered by the anodic oxide of rated thickness ≈3.4 nm (Fig. 5b) exhibits the well-defined hillocks along the whole step. The growth of the hillocks is probably related to preferred oxide growth in the regions of highest electric field (at the sites where the hillocks nucleate). Thus, the surface roughness increases after the anodic oxidation and the anodic oxide film turns out to be nonuniform in thickness. The computer analysis of the microrelief shows that the average surface roughness increases from 0.25 nm (for the nonoxidized surface) to 0.55 nm (for the oxidized surface). The maximal roughness

(from peak to valley) in the oxide is 2.6 nm. Therefore, it can be concluded that the thickness difference (thinnings) in the oxide may reach the difference between the rated thickness and maximal roughness height, i.e., 0.8 nm. This value is somewhat higher than $d_0 \approx 0.6$ nm calculated above.

The small (with allowance for film thickness nonuniformity) quantitative discrepancy between the theory and experiment can be explained by the neglect of the recombination component of the current through the diode, which is added to the thermionic overbarrier emission current and is related, for example, to recombination via surface states [15]. In this case, the expected increase in the current is even higher than that predicted by Schottky emission and the tunneling probability is lower than that obtained above (6.2×10^{-5}).

The reverse-bias portion of the *I–V* characteristic where $j \sim \sqrt{V}$ at low *V*, which corresponds to the thermal generation mechanism [12], and direct observation of electroluminescence in the forward-bias diodes are indicative of the need to consider the recombinational current component. A decrease in the barrier on hydrogen exposure probably enhances recombination, changing the shape of the *I–V* characteristic: in particular, its forward-bias branch deviates from the exponential law throughout the voltage range and both the forward-bias and reverse-bias portions of the *I–V* curve tend toward saturation as the voltage increases. Such a behavior of the *I–V* characteristic of Schottky diodes was predicted theoretically in [15] under the assumption that recombination via surface states prevails. It should be noted that the inclusion of such a recombination retains the single-slope exponential dependences of the forward and reverse current on the inverse temperature in the narrow temperature range under study and the continuous decrease in the activation energy of the current with increasing hydrogen concentration.

Since the recombination current in our diodes is significant, the determination of the semiconductor band bending with formula (1) and the density of surface states by differentiating the forward-bias branch of the *I–V* characteristic with the procedure suggested in [9] becomes difficult. Thus, hydrogen chemisorption at the AO/GaAs interface not only increases the positive charge at the interface and probability of tunneling through the oxide but also enhances the recombination activity of this interface.

Performance of the diodes as gas sensors. Figure 6 shows the dependences of ΔV_{ph} and Δj on hydrogen concentration at $T \approx 370$ K. Both ΔV_{ph} and ΔV_I increase approximately as $\sim \log P_H$ and tend toward saturation at $P_H \geq 10^{-1}$%, which is typical of such dependences [8]. The sensitivity to hydrogen (a voltage change at a given concentration) determined from the photovoltage is higher than that found from ΔV_I. Also, the determination of the hydrogen effect using ΔV_{ph} gives a higher detectivity (found by extrapolating the logarithmic part

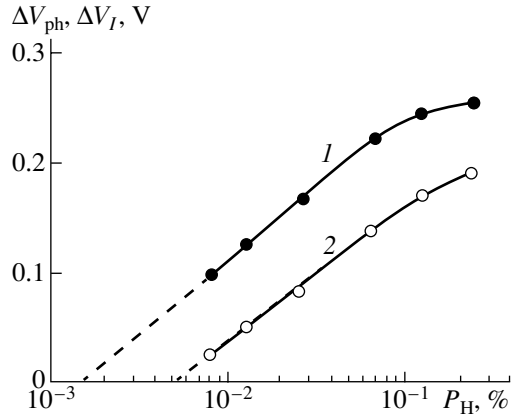


Fig. 6. Steady-state responses (*1*) ΔV_{ph} and (*2*) ΔV_I vs. P_H. $T \approx 370$ K, $d_{AO} \approx 3.4$ nm.

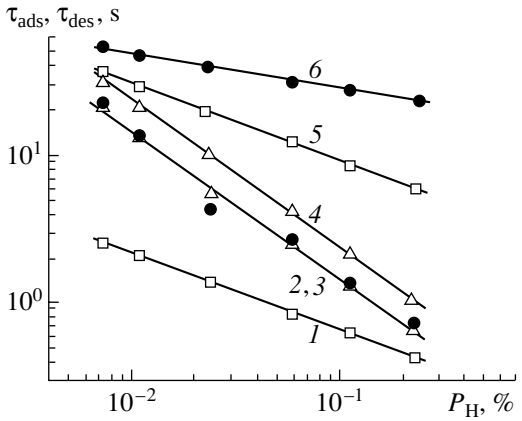


Fig. 7. (*2–4*) τ_{ads} and (*1, 5, 6*) τ_{des} vs. P_H dependences obtained from the variation of (*3, 6*) ΔV_{ph}, (*2, 5*) ΔI_f, and (*1, 4*) ΔI_r at V = (*1*) –0.3, (*2*) +0.4, and (*4, 5*) +0.3 V. $T \approx$ 370 K, $d_{AO} \approx 3.4$ nm.

of the response to $\Delta V_{ph} = 0$) than the procedure using ΔV_I. Thus, the detectivity of our Schottky diodes with respect to hydrogen is ~10^{-3}%, which is inferior to that of GaAs diodes with a native oxide (~10^{-4}%). However, as follows from the study of the hydrogen response kinetics, the speed of the diodes with AO is no less than that of the best Si and GaAs Schottky diodes [5, 8].

In general, the variation of the current and photovoltage on hydrogen exposure is nonexponential, implying heterogeneous hydrogen adsorption and desorption in the diodes with Pd [1]. Because of this, as the characteristics of the relaxation process, we take the times of adsorption, τ_{ads}, and desorption, τ_{des}, when hydrogen delivery was switched on and off, respectively. Both values were measured at a level of 0.5 in the range where the variation of the associated parameter was the fastest.

The dependences of τ_{ads} and τ_{des} on P_H at $T \approx 370$ K obtained from the variation of ΔV_{ph}, Δj_f, and Δj_r are pre-

sented in Fig. 7. In the P_H range under study, τ_{ads} and τ_{des} vary between 0.6 and 50 s, which means that the diodes are fast-responding sensors. Upon adsorption, τ_{ads} varies approximately as $\sim P_H^{-1}$ (curves *2–4*); upon desorption, τ_{des} determined from the current variation varies roughly as $\sim P_H^{-1/2}$ (curves *1*, *5*). The desorption relaxation found from the ΔV_{ph} curve (curve *6*) is the slowest. With a forward bias applied to the diode, the adsorption accelerates and the desorption slows down (cf. curves *2* and *5*). For $V < 0$, the situation is reverse (curves *1* and *3*). The maximal increase in the speed at $V < 0$ was as much as one order of magnitude (cf. curves *1* and *5*). Note that the adsorption depended on the bias relatively weakly and its rate was equally high during current and photovoltage relaxations (curves *2*, *3*).

The influence of the magnitude and sign of the electric field on hydrogen adsorption and desorption in Schottky diodes with AO can be related to the mechanism of hydrogen effect, which probably includes not only atomic hydrogen diffusion and chemisorption at the AO/GaAs interface but also the drift of protons. It is therefore natural to assume that a forward bias accelerates adsorption and a reverse bias accelerates desorption.

The activation energies of adsorption and desorption also depend on the magnitude and sign of the bias and were found to be 0.45 and 0.49 eV, respectively, at $V = +0.3$ V and 0.4 and 0.3 eV at $V = -0.3$ V.

CONCLUSION

Studying palladium/anodic oxide/gallium arsenide diodes, we found the AO layer thickness for which the sensitivity to hydrogen is maximum. The effective depth of atomic hydrogen penetration into oxide is determined. The mechanism behind the hydrogen effect involves hydrogen chemisorption, positive charge formation at the AO/GaAs interface, the drift of protons, and an enhanced probability of current carrier tunneling through the oxide. Thin tunnel-transparent anodic oxide layers are shown to be nonuniform in thickness, the surface roughness being comparable to the rated thickness. The essential role of the recombination current in Schottky diodes, which is usually neglected in the hydrogen-sensitive Si and GaAs structures, is demonstrated. It is found that the highest hydrogen detectivity is achieved from photovoltage measurements and the maximal speed of response is provided by recording the reverse current.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (grants no. 00-02-17598 and 02-02-06545).

REFERENCES

1. A. V. Evdokimov, M. N. Mushurudli, A. E. Rzhanov, *et al.*, Zarubezh. Élektron. Tekh. **2**, 12 (1988).
2. P. F. Puths, S. Ashok, S. Fovath, *et al.*, IEEE Trans. Electron Devices **28**, 1003 (1981).
3. P. E. Baguoli and A. Navvini, Solid-State Electron. **30**, 1005 (1987).
4. V. I. Gaman, P. I. Drobot, M. O. Duchenko, *et al.*, Poverkhnost, No. 11, 64 (1996).
5. S. G. Sazhin, S. V. Tikhov, V. V. Podol'skiĭ, *et al.*, Prib. Sist. Upr., No. 7, 44 (1997).
6. G. G. Kovalevskaya, N. M. Meredov, E. V. Russu, *et al.*, Zh. Tekh. Fiz. **63**, 185 (1993) [Tech. Phys. **38**, 149 (1993)].
7. H. Y. Nie and Y. Nannichi, Jpn. J. Appl. Phys. **30**, 906 (1991).
8. S. V. Tikhov, V. P. Lesnikov, V. V. Podol'skiĭ, *et al.*, Zh. Tekh. Fiz. **65** (11), 120 (1995) [Tech. Phys. **40**, 1154 (1995)].
9. V. I. Gaman, M. O. Duchenko, and V. M. Kalygina, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 1, 69 (1998).
10. S. V. Tikhov, B. I. Bednyĭ, I. A. Karpovich, *et al.*, Mikroélektronika **10**, 250 (1981).
11. L. M. Krasil'nikova, I. V. Ivonin, M. P. Yakubenya, *et al.*, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 3, 60 (1989).
12. E. H. Rhoderick, *Metal–Semiconductor Contacts* (Claredon Press, Oxford, 1978; Radio i Svyaz', Moscow, 1982).
13. I. A. Karpovich, N. V. Baidus, B. N. Zvonkov, *et al.*, Nanotechnology **12**, 425 (2001).
14. I. A. Karpovich, S. V. Tikhov, E. L. Shobolov, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **36**, 582 (2002) [Semiconductors **36**, 552 (2002)].
15. V. Kumar and W. E. Dahlke, Solid-State Electron. **20**, 143 (1977).
16. S. V. Tikhov, V. V. Martynov, I. A. Karpovich, *et al.*, Élektron. Tekh., Ser. 2: Poluprovodn. Prib. **6** (157), 18 (1982).
17. L. L. Odynets and V. M. Orlov, *Anodic Oxide Films* (Nauka, Leningrad, 1990).

Translated by M. Lebedev

Technical Physics, Vol. 48, No. 2, 2003, pp. 225–230. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 93–98.

SOLID-STATE ELECTRONICS

Diffusion Degradation Model for a Heterogeneous Photoconducting System

V. É. Bukharov, A. G. Rokakh, and S. V. Stetsyura

Chernyshevsky State University, Moskovskaya ul. 155, Saratov, 410026 Russia

e-mail: semiconductor@sgu.ssu.runnet.ru

Received May 14, 2002

Abstract—It is shown that the excitation of the electronic subsystem in heterogeneous semiconductors causes the gettering (accumulation) of defects in regions where the diffusion coefficient of defects is low. This is because this parameter in the wide-gap (active) region of the material increases, remaining low in narrow-gap areas (drains). This effect improves the radiation hardness of the heterogeneous material, as demonstrated with the CdS–PbS system, which exhibits a limited series of solid solutions. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It is known that the lifetime of II–VI (e.g., CdS) semiconductor devices, such as electron-beam-pumped lasers and photoresistors, decreases upon high-intensity irradiation because of the degradation of photoelectric parameters. The degradation of the photoconductivity manifests itself in two ways: the photoconductivity greatly decreases, while the dark conductivity significantly grows. The former effect is associated with an increase in the concentration of dislocations and point defects, which serve as centers for rapid nonradiative recombination. The nature of these centers is as yet imperfectly understood. It is generally believed that they are small associates of shallow donors, primarily Cd and Cu interstitials, with other (usually acceptor-type) defects [1, 2]. As a rule, heating to a temperature of 450 K and subsequent cooling in the dark recovers the performance of the initial (good) material.

The latter effect is usually associated with an increase in the concentration of shallow donors (primarily Cd interstitials), which can be caused by either the dissociation of donor–acceptor pairs [1] or the effect of subthreshold defect formation [3–5].

It was shown previously [6] that the stability of polycrystalline II–VI semiconducting films is improved significantly if IV–VI compounds (PbS, PbSe, and their solid solutions) are added to the initial charge. Because of the limited mutual solubility of the components, a heterogeneous material (hereafter referred to as CdS–PbS) forms. It consists of the wide-gap photoconducting matrix $Pb_xCd_{1-x}S$ and narrow-gap inclusions $Cd_yPb_{1-y}S$ (x, $y < 0.5$). The study of the stability of CdS–PbS and similar systems against degradation, as well as the construction of a mathematical model accounting for this stability, is therefore of great scientific and practical interest.

The construction of such a model is the purpose of this work. Our model utilizes the fact that irradiation variously enhances the diffusion of defects in wide- and narrow-gap areas of a heterogeneous system (CdS–PbS in particular). We generalize and extend our results published previously [7–10], review theoretical works concerned with radiation-stimulated solid diffusion, apply them to heterogeneous systems with strongly differing diffusion coefficients in their constituents, state the problem of diffusion in a heterogeneous medium, and discuss some consequences that follow from this problem.

RADIATION-STIMULATED DIFFUSION

The irradiation of solids is known to substantially enhance the diffusion of defects [11–13]. Three mechanisms of radiation-enhanced diffusion (RED) are distinguished. Vinetskiĭ and Chaĭka [11] considered the RED of recombination centers when they acquire the energy of free carriers through their recombination. In this case, the RED coefficient D^* is given by

$$D^* = D_0 \frac{n}{N_c} \exp\left(-\frac{U - \beta E_r}{kT}\right), \quad (1)$$

where D_0 is a factor weakly varying with temperature, n is the free electron concentration, N_c is the density of states in the conduction band, U is the energy barrier for the migration of the centers, E_r is the energy of an electron on the recombination level, k is the Boltzmann constant, T is temperature, and β is the empirical coefficient that equals the fraction of the free-carrier energy transferred to the center (recombination may involve intermediate levels; therefore, the energy transferred to the center may be less than E_r).

This RED mechanism takes place in the wide-gap phase of CdS–PbS, where recombination proceeds through local centers and where the energy βE_r may be significant.

1063-7842/03/4802-0225$24.00 © 2003 MAIK "Nauka/Interperiodica"

Mak [12] considered diffusion stimulated by charge exchange on a center as a result of recombination and derived the formula

$$D^* = \frac{a^2 m}{c}, \tag{2}$$

where a is the lattice parameter, c is the number of equivalent sites that can be occupied by a diffusing atom, and

$$m = \gamma_p n \left(1 + \frac{\frac{\gamma_p}{\gamma_n} n + N_c \exp\left(-\frac{E_r}{kT}\right)}{N - n + \frac{\gamma_p}{\gamma_n} N_v \exp\left(-\frac{E_g - E_r}{kT}\right)} \right)^{-1} \tag{3}$$

is the rate of charge exchange during recombination. In (3), N_v is the density of states in the valence band; γ_n and γ_p are the coefficients of trapping of electrons and holes, respectively, by a center; and E_g is the energy gap.

However, expression (2) is valid if each event of charge exchange causes a diffusion jump, i.e., if the center becomes unstable after charge exchange. Only then does the probability of the center passing jumpwise from one equilibrium state to another (adjacent) state per unit time equal m. Actually, charge exchange narrows the barrier U by only $\Delta U = 0.5$–1.0 eV [14]. With this taken into account, we refined the form of formula (2):

$$D^* = \frac{a^2 m}{c} \exp\left(-\frac{U - \Delta U}{kT}\right). \tag{4}$$

The third RED mechanism, which was suggested by Stepanov [13], relies on energy transfer to diffusing atoms when a radiation flux is elastically scattered by them. Under the assumption that the maximal energy transferred to defects from the particles is higher than the barrier U for elementary activation processes but lower than the threshold energy for atomic displacements, he derived the formula for the RED coefficient

$$D^* = D\left\{1 + \alpha \frac{E - U + kT}{E} \exp\left(\frac{U}{kT}\right)\right\}, \tag{5}$$

where D is the coefficient of equilibrium diffusion, E is the maximal energy imparted to diffusing atoms by the flux particles,

$$\alpha = \tau_t \sigma I / (1 + \tau_t \sigma I) \tag{6}$$

is the parameter of nonequilibrium, τ_t is the thermalization time of an excited atom, σ is the cross section of flux–atom elastic interaction, and I is the radiation flux.

If $E < U$, the value of D^* can be estimated with a formula similar to (5) where U in the exponent is replaced by E. Then,

$$D^* \approx D\alpha \exp\left(\frac{E}{kT}\right). \tag{7}$$

The derivation of (5) and (7) does not impose stringent restrictions on the nature of the energy gained by diffusing centers [13]. Therefore, these expressions are fairly universal. In particular, formula (7) describes RED proceeding by the first mechanism if the recombination flux I_r of electrons is substituted for the radiation flux I. Since, according to [15],

$$D = A \exp\left(-\frac{U}{kT}\right), \tag{8}$$

where A is a constant, and the energy imparted to a center through recombination $\beta E_r < U$, we have from (7)

$$D^* \approx D\alpha \exp\left(\frac{\beta E_r}{kT}\right) \approx A\tau_t \sigma_r I_r \exp\left(-\frac{U - \beta E_r}{kT}\right), \tag{9}$$

where σ_r is the cross section of charge carrier scattering by a center, $I_r = vn$ is the recombination flux, and v is the thermal velocity of electrons.

It is easy to check that (9) is equivalent to (1).

Considering a CdS–PbS photodetector as a model heterogeneous system, we estimated the coefficient of sulfur interstitial diffusion enhanced by an electron flux $I = 10^{17}$ m^{-2} s^{-1} (beam current density 10^{-2} A/m^2, accelerating voltage 2×10^4 V). None of the mechanisms mentioned above was found to noticeably enhance the diffusion of defects in narrow-gap inclusions of the CdS–PbS system (the maximal diffusion coefficient observed was $D^* = 10^{-29}$ m^2/s). Conversely, in the wide-gap component, the diffusion of defects was greatly enhanced following the first mechanism, $D^* = 10^{-14}$ m^2/s. This value appears to be quite realistic, since in experiments where CdS was irradiated by X rays (the dose rate 250 rad/s), radiolysis products diffused with a coefficient $D^* = 10^{-16}$–10^{-15} m^2/s [3], and in our case the radiation power was three orders of magnitude higher.

Such a great difference between the RED coefficients of the centers in the components of the CdS–PbS system is associated with the large energy gap and, accordingly, with a large depth of recombination levels in the wide-gap (ith) component ($E_{gi} > 2$ eV). Therefore, recombination in this component is indirect and proceeds via defect levels (Fig. 1). In narrow-gap inclusions (jth component), $E_{gj} < 0.8$ eV, recombination is direct, and defects do not acquire the energy. Thus, in an excited CdS–PbS system, defects responsible for free carrier recombination have greatly differing diffusion coefficients. The consequences of such a nonuniformity are discussed below.

MOTION OF DEFECTS IN A HETEROGENEOUS MATERIAL

Since point defects in crystals are mobile, their concentration N must obey the continuity equation

$$G - R = \mathrm{div}\mathbf{J} + \frac{\partial N}{\partial t}, \tag{10}$$

where G and R are the generation and annihilation rates, respectively. In the general case, they are functions of the coordinate $\mathbf{r}$, time t, and concentration N, and the defect flux $\mathbf{J}$ is given by

$$\mathbf{J} = -D\mathrm{grad}N + \mu N\mathbf{E}, \tag{11}$$

where D is the diffusion coefficient of defects and μ is their mobility, i.e., in the electric field $\mathbf{E}$ of a heterojunction.

In view of (11), expression (10) takes the form

$$\nabla \cdot (D\nabla N) - \nabla \cdot (\mu N\mathbf{E}) + G - R = \frac{\partial N}{\partial t}. \tag{12}$$

If the diffusion component of the flux far exceeds the drift component, which is valid, for example, in coarse grains (as in the case of CdS–PbS samples prepared by the vacuum evaporation technology described in [16]) subjected to low electric fields, one can put $\mathbf{E} = 0$ and recast (12) as

$$\nabla \cdot (D\nabla N) + G - R = \frac{\partial N}{\partial t}. \tag{13}$$

For regions with constant D, Eq. (13) can be simplified still further:

$$D\nabla^2 N + G - R = \frac{\partial N}{\partial t}. \tag{14}$$

Equation (14) is the basic equation of motion of defects in a heterogeneous semiconductor like CdS–PbS. To solve it requires initial and boundary conditions to be imposed.

Since the concentration of defects grows during irradiation and much exceeds the initial concentration N_0 (which causes degradation), one can put $N_0 = 0$; then, the initial condition has the form

$$N(\mathbf{r}, 0) = 0. \tag{15}$$

For joining together the solutions at the interface, we direct grad$N$ along the x axis. Then, Eq. (11) can be written in scalar form, where the operator ∇ is replaced by differentiation with respect to x. In this case, the diffusion component J_{dif} of the flux is

$$J_{\mathrm{dif}} = -D(\partial N/\partial x). \tag{16}$$

On the other hand, defects can be treated in terms of the molecular kinetic theory (MKT) [14]. Let some surface S_{ij}, say, the plane $x = l$, separate the defect area i ($x < l$) and defect-free area j ($x > l$). Then, using the MKT, one finds that the diffusion flux J_{dif} of defects at the interface is $J_{\mathrm{dif}} = N(l)v_D/6$, where $v_D = a/\tau$ is the defect velocity in the crystal (a is the lattice constant, which plays the role of the free path when considering the diffusion of defects, and τ is the mean residence time of a defect in an equilibrium position). Expressing v_D through D [14],

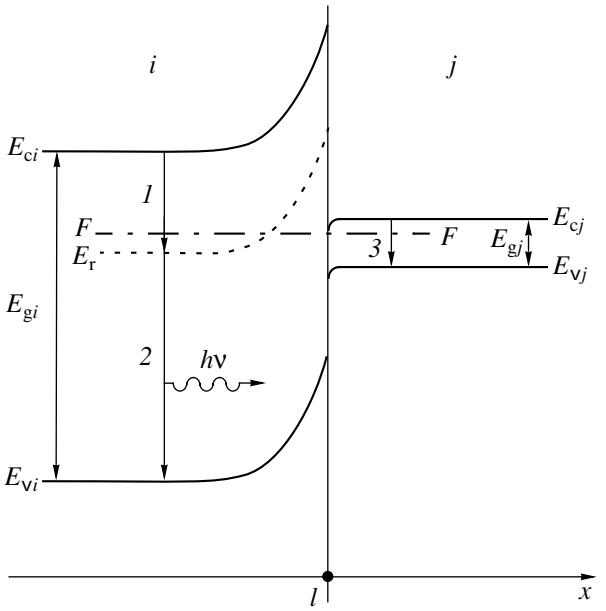


Fig. 1. Band diagram of the heterogeneous photoconducting system. E_{ci} and E_{cj}, conduction band bottoms; E_{vi} and E_{vj}, valence band tops; F, Fermi level; E_r, recombination (defect) level; E_{gi} and E_{gj}, energy gaps; *1*, capture of an electron from the conduction band to a defect-related recombination level; *2*, recombination of the captured electron, photon release, and energy transfer to the defect; and *3*, direct (band-to-band) recombination.

$$v_D = 6D/a, \tag{17}$$

we find that $J = DN(l)/a$.

Now if the defect concentration N is $N(l - 0) = N_i$ to the left of the point $x = l$ and $N(l + 0) = N_j$ to the right of this point (Fig. 2), the diffusion flux J_{dif} across the plane $x = l$ is given by the difference $J_i - J_j$:

$$\begin{aligned} J_{\mathrm{dif}} &= J_i - J_j \\ &= [D(l-0)N(l-0) - D(l+0)N(l+0)]/a. \end{aligned} \tag{18}$$

Assuming that $D(x)$ and $N(x)$ are continuous functions and taking into account that the least distance dx in a crystal that has a physical meaning is the lattice parameter a, we obtain formula (16) from (18). If at least one of these functions experiences a discontinuity at the point $x = l$, we obtain the joining condition at the interface from (18). If $D(x)$ and $N(x)$ are piecewise continuous functions and everywhere have derivatives on the left and right and if the flux is continuous at the

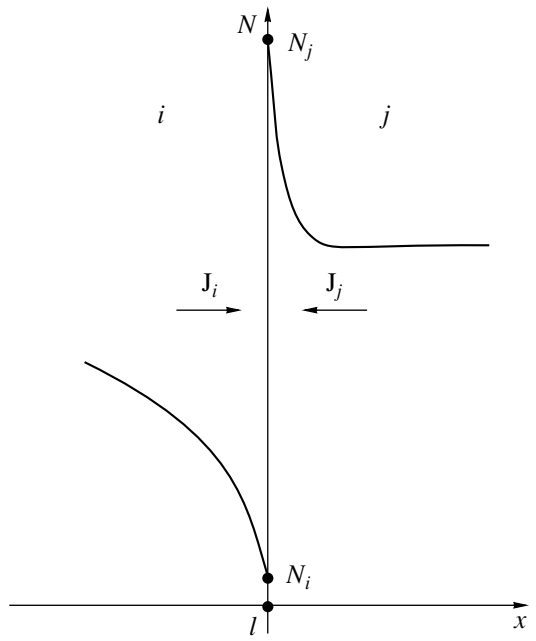


Fig. 2. Defect concentration profile and defect fluxes at the interface between the wide (*i*)- and narrow (*j*)-gap phases of the heterogeneous semiconductor. *N*, defect concentration; *x*, coordinate; *l*, coordinate of the interface; and $\mathbf{J}_i$ and $\mathbf{J}_j$, defect fluxes at the interface.

interface (defects do not annihilate at the interface), the joining condition appears as

$$-D_i\frac{\partial N_i}{\partial x}\bigg|_{x=l} = -D_j\frac{\partial N_j}{\partial x}\bigg|_{x=l} = \frac{D_i}{a_i}N_i(l) - \frac{D_j}{a_j}N_j(l). \quad (19)$$

Since grad*N* is aligned with the *x* axis, (19) can be recast in vector form

$$-D_i|\nabla N_i|\big|_{S_{ij}} = -D_j|\nabla N_j|\big|_{S_{ij}} = [(D_i/a_i)N_i - (D_j/a_j)N_j]\big|_{S_{ij}}. \quad (20)$$

Here, the interface between the regions *i* and *j* may be an arbitrary (including closed) surface and not necessarily a plane.

Note that if $N_i(l) \approx N_j(l)$, $a_i \approx a_j$, and $D_i \gg D_j$, the right-hand side of (19) is positive and the flux of defects from the region *i* to the region *j* arises; in other words, defects are gettered by the region with a smaller diffusion coefficient (as if stuck in it). Thus, the difference in the RED rate in the narrow- and wide-gap regions of a heterogeneous photoconducting system (such as CdS–PbS) causes the gettering of defects by narrow-gap inclusions.

If the regions with a smaller diffusion coefficient are limited (as takes place in CdS–PbS), the defect flux mentioned above causes local minima of $N(\mathbf{r})$ to appear inside narrow-gap inclusions. Let the number of these minimum points generate a set ρ (in the simplest case, ρ = 1). Then, in the general case, the boundary condition for a narrow-gap inclusion has the form

$$\mathrm{grad}N|_{\mathbf{r}\in\rho} = 0. \quad (21)$$

For the wide-gap region, the boundary condition is the same with the only difference that here points of extremum are maximum points. Therefore, since this region can be considered as unbounded (its size is much greater than that of a narrow-gap inclusion), here ρ represents, as a rule, a surface (or a set of surfaces).

To summarize, we mathematically state the problem of defect migration in the CdS–PbS heterogeneous photoconductor in the general form

$$\nabla\cdot(D_i\nabla N_i) - \nabla\cdot(\mu_i N_i\mathbf{E}_i) + G_i - R_i = \partial N_i/\partial t, \quad (22)$$

$$\nabla\cdot(D_j\nabla N_j) - \nabla\cdot(\mu_j N_j\mathbf{E}_j) + G_j - R_j = \partial N_j/\partial t, \quad (23)$$

$$N_i(\mathbf{r}, 0) = 0, \quad (24)$$

$$N_j(\mathbf{r}, 0) = 0, \quad (25)$$

$$\mathrm{grad}N_i|_{\mathbf{r}\in\rho_i} = 0, \quad (26)$$

$$\mathrm{grad}N_j|_{\mathbf{r}\in\rho_j} = 0, \quad (27)$$

$$-D_i|\nabla N_i|\big|_{S_{ij}} = -D_j|\nabla N_j|\big|_{S_{ij}} = [(D_i/a_i)N_i - (D_j/a_j)N_j]\big|_{S_{ij}}. \quad (28)$$

Recall that the subscripts *i* and *j* refer to the wide- and narrow-gap regions, respectively. Expression (28), which is the joining condition at the interface, refers to adjacent regions separated by the surfaces S_{ij}.

Since the only property of the CdS–PbS heterogeneous system used in stating the problem is the nonuniform distribution of the diffusion coefficient, our considerations are applicable to any solid-state heterogeneous system featuring the nonuniform distribution of the diffusion coefficient.

Below, we introduce some notions and discuss several results following from boundary-value problem (22)–(28).

MEAN CONCENTRATIONS AND MEAN RATE OF DEFECT GENERATION IN A HETEROGENEOUS MEDIUM

For any regions of a heterogeneous photoconductor, one can introduce the parameter G_{eff} that has the meaning of the effective mean rate of defect formation (by analogy with a homogeneous photoconductor):

$$G_{\mathrm{eff}} = \langle\partial N/\partial t + R\rangle, \quad (29)$$

where the angular brackets mean a space average.

We assume that the diffusion migration of defects is the only reason for the effects discussed above; that is,

Eq. (13) is the basic one. In view of this equation, let us recast (29) as

$$G_{\text{eff}} = \frac{1}{V}\int_V [\nabla \cdot (D\nabla N) + G]dV = \langle G\rangle + \frac{1}{V}\oint_S D\,\text{grad}\,N dS. \quad (30)$$

Here, V is the volume of the corresponding space and S is the boundary surface. In (30), we used the divergence theorem. Since, as was shown above, the defect flux across S is directed from the wide- to narrow-gap region, the integral $\oint_S \text{grad}\,N\,dS$ is positive for a narrow-gap inclusion and negative for the wide-gap matrix. Hence, the effective rate of defect generation in the wide-gap phase is lower and in the narrow-gap phase is higher than the generation rate $\langle G\rangle$ in the corresponding homogeneous material. In other words, the presence of the narrow-gap phase, where the radiation-induced enhancement of defect diffusion is lower than in the wide-gap phase, decreases the effective rate of defect generation in the wide-gap phase of a heterogeneous semiconductor.

Let $\bar{N}_\omega$ ($\omega = i, j$) be the mean concentrations in the wide- and narrow-gap regions:

$$\bar{N}_\omega = \langle N_\omega\rangle = \frac{1}{V_\omega}\int_{V_\omega} N_\omega dV, \quad (31)$$

where V_ω is the volume of the corresponding space.

Expressing the recombination term as $R = \kappa N^2$ (κ is the recombination coefficient), which is valid for the direct recombination of Frenkel defects, and integrating (13) over the volume V_i of the wide-gap region, we arrive at an equation for $\bar{N}_i$:

$$V_i\frac{d\bar{N}_i}{dt} = V_i\langle G_i\rangle - \kappa\int_{V_i} N_i^2 dV + \oint_{S_i} D_i\nabla N_i dS, \quad (32)$$

where S_i is the surface bounding the wide-gap region.

Since the variance $\langle(N - \bar{N})^2\rangle = \langle N^2\rangle - \langle N\rangle^2$ is nonnegative, we have

$$V_i\langle N_i^2\rangle = \int_{V_i} N_i^2 dV \geq \frac{1}{V_i}\left(\int_{V_i} N dV\right)^2 = V_i\bar{N}_i^2. \quad (33)$$

Consequently, in view of (32) and (33), we can write

$$\frac{d\bar{N}_i}{dt} \leq \langle G_i\rangle - \kappa\bar{N}_i^2 + \frac{1}{V_i}\oint_S S_i\nabla N_i dS \quad (34)$$

and conclude that $\tilde{N}_i$, which obeys the equation

$$\frac{d\tilde{N}_i}{dt} = \langle G_i\rangle - \kappa\tilde{N}_i^2 + \frac{1}{V_i}\oint_S D_i\nabla N_i dS, \quad (35)$$

bounds $\bar{N}_i$ from above; that is, $\bar{N}_i \leq \tilde{N}_i$. Since the two last terms in (35) are nothing but G_{eff}, which, in the wide-gap region, is below the mean generation rate $\langle G\rangle$, we find that $\bar{N}_i < \tilde{N}_i < N_0$, where N_0 is the defect concentration in a homogeneous material at the same defect generation rate G. This means that the concentration of radiation-induced defects in the wide-gap phase of a heterogeneous semiconductor is always lower than in a homogeneous system of the same composition; hence, the wide-gap phase is superior to the homogeneous material in radiation hardness.

The presence of an electric field may change the defect distribution up to the point where $\bar{N}_i > N_0$; however, in this case, too, $\bar{N}_i$ will be lower than the concentration of defects under the condition when the coefficient of defect diffusion is the same throughout the material.

When implementing this mechanism of increasing the radiation hardness, one should bear in mind that the mere difference in D is insufficient for this mechanism to be efficient. It is also necessary that the diffusion coefficient in the active region (the wide-gap region in the case of CdS–PbS) be so high that defects can reach the gettering zone for the time of enhanced diffusion (i.e., during irradiation). Our calculations and experiments with CdS [3, 6, 8] show that such a situation occurs in an irradiated heterogeneous CdS–PbS semiconductor.

CONCLUSIONS

The basic results of this work are as follows.

(1) A different degree of defect diffusion enhancement in the wide-gap matrix and narrow-gap inclusions of a heterogeneous photoconductor improves its radiation hardness in comparison with a homogeneous photoconductor provided that the diffusion coefficient in the active region of the heterogeneous material is sufficiently high.

(2) If the RED coefficients in the wide- and narrow-gap regions of heterogeneous systems differ substantially, radiation-induced and processing-related defects are gettered by the lower coefficient region. This decreases greatly the mean concentration of defects and the mean effective rate of their generation (formula (29)) in the higher coefficient region.

(3) These processes may be responsible for the experimentally observed fact that the radiation hardness of an CdS heterogeneous semiconductor is improved by adding PbS.

REFERENCES

1. N. E. Korsunshaya, I. V. Markevich, B. R. Dzhumaev, *et al.*, Semicond. Phys. Quantum Electron. Optoelectron. **2** (1), 42 (1999).
2. É. N. Zamanov and M. A. Dzhafarov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **29**, 1411 (1995) [Semiconductors **29**, 732 (1995)].
3. V. S. Vavilov, A. E. Kiv, and O. R. Niyazova, *Mechanisms of Defect Formation and Migration in Semiconductors* (Nauka, Moscow, 1981).
4. M. S. Yunusov, S. N. Abdurakhmanova, M. A. Zaĭtsovskaya, *et al.*, *Subthreshold Radiation Effects in Semiconductors* (Fan, Tashkent, 1989).
5. V. L. Vinetskiĭ and G. A. Kholodar', *Statistical Interaction of Electrons and Defects in Semiconductors* (Naukova Dumka, Kiev, 1969).
6. A. G. Rokakh, S. V. Stetsyura, N. B. Trofimova, *et al.*, Neorg. Mater. **35**, 552 (1999).
7. A. G. Rokakh, Pis'ma Zh. Tekh. Fiz. **10**, 820 (1984) [Sov. Tech. Phys. Lett. **10**, 344 (1984)].
8. V. É. Bukharov, A. G. Rokakh, and S. V. Stetsyura, Pis'ma Zh. Tekh. Fiz. **25** (3), 66 (1999) [Tech. Phys. Lett. **25**, 111 (1999)].
9. V. É. Bukharov and A. G. Rokakh, Pis'ma Zh. Tekh. Fiz. **25** (24), 55 (1999) [Tech. Phys. Lett. **25**, 994 (1999)].
10. A. G. Rokakh and N. B. Trofimova, Zh. Tekh. Fiz. **71** (7), 140 (2001) [Tech. Phys. **46**, 926 (2001)].
11. V. L. Vinetskiĭ and G. E. Chaĭka, Fiz. Tverd. Tela (Leningrad) **24**, 2170 (1982) [Sov. Phys. Solid State **24**, 1236 (1982)].
12. V. T. Mak, Zh. Tekh. Fiz. **63** (3), 173 (1993) [Tech. Phys. **38**, 246 (1993)].
13. V. A. Stepanov, Zh. Tekh. Fiz. **68** (8), 67 (1998) [Tech. Phys. **43**, 938 (1998)].
14. B. I. Boltaks, *Diffusion and Point Defects in Semiconductors* (Nauka, Leningrad, 1972).
15. H. H. Woodbury, in *Physics and Chemistry of II–VI Compounds*, Ed. by M. Aven and J. S. Prener (North-Holland, Amsterdam, 1967; Mir, Moscow, 1970).
16. A. G. Rokakh, A. V. Kumakov, and N. V. Elagina, RF Patent No. 845,685 (1993); Byull. Izobret., No. 25 (1993).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 231–238. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 99–106.

ACOUSTIC, ACOUSTOELECTRONICS

Theory of an SAW Filter Operating in Weakly Coupled Resonance Modes

V. F. Dmitriev

Avangard–Elionika Closed Joint-Stock Company, St. Petersburg, 195271 Russia

e-mail: elionica@rol.ru

Received June 21, 2002

Abstract—Modified coupled-mode equations are employed to develop a theory of filters that use waveguide modes of two surface acoustic wave (SAW) resonators weakly coupled in the transverse direction. Results of simulation based on this theory are compared with the earlier experiment.© *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The coupled mode (COM) method is an efficient technique for calculating parameters of SAW filters. The conventional COM method (see, e.g., [1]), which relies on a system of inhomogeneous differential equations, involves too complicated a procedure for designing SAW devices with arbitrarily varying parameters. In the framework of this theory, it is difficult to allow for such factors as the variable period of the structure, the arbitrary polarity of the electrodes connected to the contact lines, apodization, and the nonuniform charge distribution over the electrodes. All these factors are easy to take into account by the method that is based on modified equations for coupled modes and deals with an elementary unit of the structure. This method also allows one to use more complex models of the structure.

The method of modified equations has been used to simulate various SAW devices [2–4]. In this paper, we elaborate a theory of a narrow-band SAW filter that is most promising for communications. These filters employ waveguide modes in a structure of two SAW resonators weakly coupled in the transverse direction (transversely coupled resonator filters, TCRFs). TCRFs attract researchers' attention because they feature a low insertion loss and high out-of-band suppression. Earlier (see, e.g., review [5]), TCRFs were analyzed based on the conventional coupled mode theory (COM theory). However, TCRFs with variable polarity of electrode pairs connected to the contact buses are difficult to simulate in terms of the conventional COM theory. In the TCRFs considered in this paper, the polarity of the electrodes is chosen so that at least two longitudinal resonance modes are excited. Note that, owing to both the weak transverse coupling between the two resonators and the waveguide effect, either longitudinal mode splits into two modes. Thus, in TCRFs with variable polarity of electrode pairs connected to the contact bus, four or even more modes can be excited simultaneously. A four-mode filter provides a wider passband and higher shape factor than a usual two-mode TCRF. The theory proposed below can be used for simulating conventional TCRFs, as well as TCRFs with variable electrode-pair polarity.

In this paper, a theory of filters that use waveguide modes in two SAW resonators with weak transverse coupling is developed. Calculations based on this theory are compared with the earlier experiment [6, 7].

TRANSVERSE ACOUSTIC FIELD DISTRIBUTION IN A TCRF

Let $R(y, z, \omega)$ and $S(y, z, \omega)$ be two coupled inhomogeneous plane waves of angular frequency ω that propagate in a semi-infinite piezoelectric with interdigital transducers (IDT-1 and IDT-2) and reflecting arrays (RA1-L, RA1-R, RA2-L, and RA2-R) in the form of metal strips applied on its surface (Fig. 1). The wave $R(y, z, \omega)$ travels in the positive z direction; the wave $S(y, z, \omega)$, in the negative z direction. We will also assume that the transverse field distribution is z independent and that the structure is limited in the y (trans-

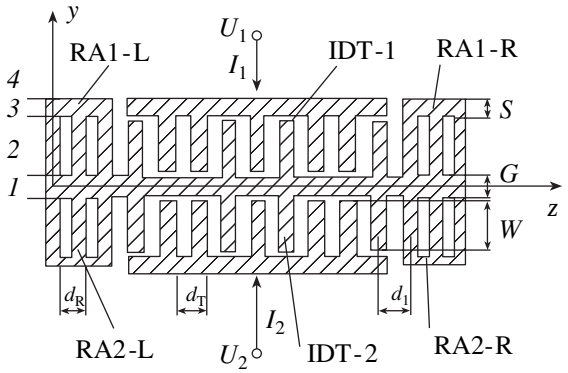


Fig. 1. Layout of a TCRF.

1063-7842/03/4802-0231$24.00 © 2003 MAIK "Nauka/Interperiodica"

verse) direction; therefore, waveguide modes can exist in this structure. The inhomogeneous plane waves can be written as

$$R(y, z, \omega) = R(\omega)\Psi(y)\exp(-jkz), \quad (1)$$

$$S(y, z, \omega) = S(\omega)\Psi(y)\exp(+jkz), \quad (2)$$

where $R(\omega)$ and $S(\omega)$ are the complex amplitudes of the waves, $\Psi(y)$ is the transverse acoustic field distribution, and k is the wave number.

We will seek the field distribution $\Psi(y)$ as a solution to the wave equation for the complex amplitudes of waves traveling in the positive z direction:

$$\frac{d^2\Psi^{(t)}(y)}{d^2y} + k_t^2 d\Psi^{(t)}(y) = 0, \quad (3)$$

where $t = 1, 2, 3$, or 4 is the number of the structure's region (Fig. 1) for which the equation is written.

The acoustic field distribution must satisfy the boundary conditions

$$\Psi^{(1)}(y) = \Psi^{(2)}(y) \quad \text{at} \quad y = G/2, \quad (4)$$

$$\Psi^{(2)}(y) = \Psi^{(3)}(y) \quad \text{at} \quad y = G/2 + W, \quad (5)$$

$$\Psi^{(3)}(y) = \Psi^{(4)}(y) \quad \text{at} \quad y = G/2 + W + S, \quad (6)$$

$$\frac{\partial\Psi^{(1)}(y)}{\partial y} = \frac{\partial\Psi^{(2)}(y)}{\partial y} \quad \text{at} \quad y = G/2, \quad (7)$$

$$\frac{\partial\Psi^{(2)}(y)}{\partial y} = \frac{\partial\Psi^{(3)}(y)}{\partial y} \quad \text{at} \quad y = G/2 + W, \quad (8)$$

$$\frac{\partial\Psi^{(3)}(y)}{\partial y} = \frac{\partial\Psi^{(4)}(y)}{\partial y} \quad \text{at} \quad y = G/2 + W + S. \quad (9)$$

Since the structure is symmetric about the z axis, a solution to Eq. (3) under boundary conditions (4)–(9) will be sought in the region $y > 0$ as the expansion in orthonormal set of functions

$$\Psi^{(t)}(y) = \sum_n \Psi_n^{(t)}(y), \quad (10)$$

where $\Psi_n(y) = A_n\sin(k_{tn}y) + B_n\cos(k_{tn}y)$.

In region *1*, $k_{tn}^2 = -k_{1n}^2 = -(k_n^2 - k_m^2)$; in region *2*, $k_{tn}^2 = k_{2n}^2 = k_e^2 - k_n^2$; in region *3*, $k_{3n} = k_{1n}$; and in region *4*, $k_{tn}^2 = -k_{4n}^2 = -(k_n^2 - k_0^2)$. Here, k_m, k_e, and k_0 are the SAW wave numbers for an infinite piezoelectric covered by a continuous metal film, covered by an infinitely long array of infinitely long electrodes, and uncovered, respectively, and k_n is the wave number of an nth mode. In the general case, waveguide modes of n ($n = 0, 1, 2, 3, \ldots$) types can exist in the structure, each having a particular transverse field distribution. Therefore, we will seek a solution for each of them.

Taking into account the sign of the propagation constant k_{tn} and the symmetry of the structure, we look for a solution to Eq. (3) in regions *1–4* in the following form:

in region *1* ($0 < y < G/2$),

$$\Psi_n^{(1)}(y) = B_{1n}F_B,$$

where

$$F_B = \begin{cases} \cosh(k_{1n}G/2) & \text{for symmetric modes} \\ \sinh(k_{1n}G/2) & \text{for antisymmetric modes;} \end{cases} \quad (11)$$

in region *2* ($G/2 < y < W + G/2$),

$$\Psi_n^{(2)}(y) = A_{2n}\sin(k_{2n}y) + B_{2n}\cos(k_{2n}y); \quad (12)$$

in region *3* ($W + G/2 < y < W + G/2 + S$),

$$\Psi_n^{(3)}(y) = B_{3n}\exp(-k_{3n}y); \quad (13)$$

and in region *4* ($y > W + G/2 + S$),

$$\Psi_n^{(4)}(y) = B_{4n}\exp(-k_{4n}y). \quad (14)$$

The coefficients $B_{1n}, A_{2n}, B_{2n}, A_{3n}, B_{3n}$, and B_{4n} can be found from boundary conditions (4)–(9) and the orthonormality condition

$$\int_{-\infty}^{\infty} dy\Psi_n^{(t)}(y)\Psi_m^{(t)}(y) = \delta_{nm}W, \quad (15)$$

where δ_{nm} is the Kronecker symbol.

Using boundary conditions (5) and (8) and expressions (12) and (13), a dispersion relation that relates the wave numbers of acoustic waveguide modes to their angular frequency can be derived. The existence conditions for a solution to the dispersion relation are discussed in [1].

MODIFIED COM EQUATIONS FOR A TCRF

To calculate the TCRF's input admittance, let us use Eqs. (9)–(11) from [3] and expand the acoustic wave field in waveguide modes. Then, with expressions (1), (2), and (10), we can write

$$R(y, z, \omega) = \sum_n R_n(\omega)\exp(-jk_nz)\Psi_n(y), \quad (16)$$

$$S(y, z, \omega) = \sum_n S_n(\omega)\exp(+jk_nz)\Psi_n(y). \quad (17)$$

Next, we perform the following standard procedure. Substituting expansions (16) and (17) into Eqs. (9) and (10) from [3], multiplying both sides of the equations by $\Psi_n(y)$, and integrating the result with respect to y

from $-\infty$ to $+\infty$, we obtain, in view of orthonormality condition (15),

$$S_{K,n}(\omega) = r_K\eta_{1,K}\exp[-j\kappa_n p_K]R_{K,n}(\omega) + \eta_{1,K}(1-|r_K|^2)^{1/2}\exp[-j\kappa_n p_K]S_{K+1,n}(\omega) + \xi_K(\omega)\eta_{2,K}\exp[-j\kappa_n p_K/2](T_n^{(1)}U_1 + T_n^{(2)}U_2), \quad (18)$$

$$R_{K+1,n}(\omega) = \eta_{1,K}(1-|r_K|^2)^{1/2}\exp[-j\kappa_n p_K]R_{K,n}(\omega) + r_K\eta_{1,K}\exp[-j\kappa_n p_K]S_{K+1,n}(\omega) + \xi_K(\omega)\eta_{2,K}\exp[-j\kappa_n p_K/2](T_n^{(1)}U_1 + T_n^{(2)}U_2), \quad (19)$$

where r_K is the complex reflection coefficient of a Kth electrode, $\kappa_n = k_n - k_0 - j\alpha$, $k_0 = 2\pi/p_K$, p_K is the center distance between $(K+1)$th and Kth electrodes, ξ_K is the coefficient of SAW transformation by a Kth electrode [3], $\eta_{1,K} = W_{1,K}/W$, $\eta_{2,K} = W_{2,K}/W$, W is the maximal aperture, $W_{1,K}$ is the overlap of adjacent electrodes, $W_{2,K} = W$ when dummy electrodes are used and $W_{2,K} = W_{1,K}$ otherwise, α is the total SAW propagation loss in the electrode structure per unit length, and L_K is the width of a Kth electrode. The SAW reflection (transformation) center is assumed to be at the center of each electrode.

The quantities $T_n^{(1)}$ and $T_n^{(2)}$ are defined as

$$T_n^{(1)} = (1/W)\int_{G/2}^{G/2+W} dy\Psi_n(y),$$

$$T_n^{(2)} = (1/W)\int_{-(G/2+W)}^{-G/2} dy\Psi_n(y).$$

Note that $T_n^{(1)} = T_n^{(2)}$ for symmetric waves and $T_n^{(1)} = -T_n^{(2)}$ for antisymmetric ones. Let the number of electrodes in either IDT and in each of the reflecting arrays be N_T and N_R, respectively. Next, we write Eq. (11) from [3] for either transducer, substitute expansions (16) and (17) into the resulting equations, multiply both their sides by $\Psi_n(y)$, and integrate the result over the aperture of either transducer to arrive at formulas for the variation of the current in the IDTs due to the transformation of the forward- and backward-traveling waves:

$$I_{1,K,n}(\omega) - I_{1,K+1,n}(\omega) = 2\xi_K(\omega)\exp[-j\kappa_n p_K/2] \times T_n^{(1)}[R_{K,n}(\omega) + S_{K,n}(\omega)], \quad (20)$$

$$I_{2,K,n}(\omega) - I_{2,K+1,n}(\omega) = 2\xi_K(\omega)\exp[-j\kappa_n p_K/2] \times T_n^{(2)}[R_{K,n}(\omega) + S_{K,n}(\omega)]. \quad (21)$$

The change in the current in the buses that is caused by the voltage drop across the IDT capacitance will be taken into account later.

Relationships (18)–(21) can be written in matrix form:

$$\begin{vmatrix} S_{K,n}(\omega) \\ R_{K+1,n}(\omega) \\ \Delta I_{1,K,n}(\omega) \\ \Delta I_{2,K,n}(\omega) \end{vmatrix} = \begin{vmatrix} P(1,1) & P(1,2) & P(1,3) & P(1,4) \\ P(2,1) & P(2,2) & P(2,3) & P(2,4) \\ P(3,1) & P(3,2) & P(3,3) & P(3,4) \\ P(4,1) & P(4,2) & P(4,3) & P(4,4) \end{vmatrix} \times \begin{vmatrix} R_{K,n}(\omega) \\ S_{K+1,n}(\omega) \\ U_1 \\ U_2 \end{vmatrix}, \quad (22)$$

where $\Delta I_{1,K,n}(\omega) = I_{1,K,n}(\omega) - I_{1,K+1,n}(\omega)$, $\Delta I_{2,K,n}(\omega) = I_{2,K,n}(\omega) - I_{2,K+1,n}(\omega)$,

$$P(1,1) = r_K\eta_{1,K}\exp[-j\kappa_n p_K],$$

$$P(1,2) = \eta_{1,K}(1-|r_K|^2)^{1/2}\exp[-j\kappa_n p_K],$$

$$P(1,3) = \xi_K(\omega)\eta_{2,K}\exp[-j\kappa_n p_K/2]T_n^{(1)},$$

$$P(1,4) = \xi_K(\omega)\eta_{2,K}\exp[-j\kappa_n p_K/2]T_n^{(2)},$$

$$P(2,1) = \eta_{1,K}(1-|r_K|^2)^{1/2}\exp[-j\kappa_n p_K],$$

$$P(2,2) = r_K\eta_{1,K}\exp[-j\kappa_n p_K],$$

$$P(2,3) = \xi_K(\omega)\eta_{2,K}\exp[-j\kappa_n p_K/2]T_n^{(1)},$$

$$P(2,4) = \xi_K(\omega)\eta_{2,K}\exp[-k\kappa_n p_K/2]T_n^{(2)},$$

$$P(3,1) = P(3,2) = 2\xi_K(\omega)\exp[-j\kappa_K/2]T_n^{(1)},$$

$$P(4,1) = P(4,2) = 2\xi_K(\omega)\exp[-j\kappa_n P_K/2]T_n^{(2)},$$

$$P(3,3) = P(3,4) = P(4,3) = P(4,4) = 0.$$

To calculate the components of the TCRF's P matrix, we will first find the components of the total P matrix for two SAW structures (with matrices $[P^{(1)}]$ and $[P^{(2)}]$) connected in series. For the sake of simplicity, we omit the subscript n from relationships that follow. Then, the $[P^{(1)}]$ matrix is given by

$$\begin{vmatrix} S_K(\omega) \\ R_{K+1}(\omega) \\ \Delta I_{1,K}(\omega) \\ \Delta I_{2,K}(\omega) \end{vmatrix} = \begin{vmatrix} P^{(1)}(1,1) & P^{(1)}(1,2) & P^{(1)}(1,3) & P^{(1)}(1,4) \\ P^{(1)}(2,1) & P^{(1)}(2,2) & P^{(1)}(2,3) & P^{(1)}(2,4) \\ P^{(1)}(3,1) & P^{(1)}(3,2) & P^{(1)}(3,3) & P^{(1)}(3,4) \\ P^{(1)}(4,1) & P^{(1)}(4,2) & P^{(1)}(4,3) & P^{(1)}(4,4) \end{vmatrix} \begin{vmatrix} R_K(\omega) \\ S_{K+1}(\omega) \\ U_1 \\ U_2 \end{vmatrix}, \quad (23)$$

where $\Delta I_{1,K}(\omega) = I_{1,K}(\omega) - I_{1,K+1}(\omega)$ and $\Delta I_{2,K}(\omega) = I_{2,K}(\omega) - I_{2,K+1}(\omega)$.

The $[P^{(2)}]$ matrix will be defined as

$$\begin{vmatrix} S_{K+1}(\omega) \\ R_{K+1}(\omega) \\ \Delta I_{1,K+1}(\omega) \\ \Delta I_{2,K+1}(\omega) \end{vmatrix} = \begin{vmatrix} P^{(2)}(1,1) & P^{(2)}(1,2) & P^{(2)}(1,3) & P^{(2)}(1,4) \\ P^{(2)}(2,1) & P^{(2)}(2,2) & P^{(2)}(2,3) & P^{(2)}(2,4) \\ P^{(2)}(3,1) & P^{(2)}(3,2) & P^{(2)}(3,3) & P^{(2)}(3,4) \\ P^{(2)}(4,1) & P^{(2)}(4,2) & P^{(2)}(4,3) & P^{(2)}(4,4) \end{vmatrix} \begin{vmatrix} R_{K+1}(\omega) \\ S_{K+2}(\omega) \\ U_1 \\ U_2 \end{vmatrix}, \quad (24)$$

where $\Delta I_{1,K+1}(\omega) = I_{1,K+1}(\omega) - I_{1,K+2}(\omega)$ and $\Delta I_{2,K+1}(\omega) = I_{2,K+1}(\omega) - I_{2,K+2}(\omega)$.

Using systems of equations (23) and (24) with arbitrary coefficients, which are written for two series-connected SAW structures of the TCRF, we easily obtain the components of the total P matrix:

$$\begin{vmatrix} S_K(\omega) \\ R_{K+2}(\omega) \\ \Delta I_1(\omega) \\ \Delta I_2(\omega) \end{vmatrix} = \begin{vmatrix} P^{(S4)}(1,1) & P^{(S4)}(1,2) & P^{(S4)}(1,3) & P^{(S4)}(1,4) \\ P^{(S4)}(2,1) & P^{(S4)}(2,2) & P^{(S4)}(2,3) & P^{(S4)}(2,4) \\ P^{(S4)}(3,1) & P^{(S4)}(3,2) & P^{(S4)}(3,3) & P^{(S4)}(3,4) \\ P^{(S4)}(4,1) & P^{(S4)}(4,2) & P^{(S4)}(4,3) & P^{(S4)}(4,4) \end{vmatrix} \begin{vmatrix} R_K(\omega) \\ S_{K+2}(\omega) \\ U_1 \\ U_2 \end{vmatrix}, \quad (25)$$

where $\Delta I_1(\omega) = I_{1,K}(\omega) - I_{1,K+2}(\omega)$, $\Delta I_2(\omega) = I_{2,K}(\omega) - I_{2,K+2}(\omega)$,

$$P^{(S4)}(1,1) = P^{(1)}(1,1) + P^{(1)}(1,2)P^{(2)}(1,1)P^{(1)}(2,1)/P_0; \quad (26)$$

$$P^{(S4)}(1,2) = P^{(1)}(1,2)P^{(2)}(1,2)/P_0; \quad (27)$$

$$P^{(S4)}(1,3) = P^{(1)}(1,3) + P^{(1)}(1,2)[P^{(2)}(1,3) + P^{(2)}(1,1)P^{(1)}(2,3)]/P_0; \quad (28)$$

$$P^{(S4)}(1,4) = P^{(1)}(1,4) + P^{(1)}(1,2)[P^{(2)}(1,4) + P^{(2)}(1,1)P^{(1)}(2,4)]/P_0; \quad (29)$$

$$P^{(S4)}(2,1) = P^{(1)}(2,1)P^{(2)}(2,1)/P_0; \quad (30)$$

$$P^{(S4)}(2,2) = P^{(2)}(2,2) + P^{(2)}(2,1)P^{(1)}(2,2)P^{(2)}(1,2)/P_0; \quad (31)$$

$$P^{(S4)}(2,3) = P^{(2)}(2,3) + P^{(2)}(2,1)[P^{(1)}(2,3) + P^{(2)}(1,3)P^{(1)}(2,2)]/P_0; \quad (32)$$

$$P^{(S4)}(2,4) = P^{(2)}(2,4) + P^{(2)}(2,1)[P^{(1)}(2,4) + P^{(2)}(1,4)P^{(1)}(2,2)]/P_0; \quad (33)$$

$$P^{(S4)}(3,1) = P^{(1)}(3,1) + P^{(1)}(2,1)[P^{(2)}(3,1) + P^{(2)}(1,1)P^{(1)}(3,2)]/P_0; \quad (34)$$

$$P^{(S4)}(3,2) = P^{(2)}(3,2) + P^{(2)}(1,2)[P^{(1)}(3,2) + P^{(1)}(2,2)P^{(2)}(3,1)]/P_0; \quad (35)$$

$$P^{(S4)}(3,3) = P^{(1)}(3,3) + P^{(2)}(3,3) + \{P^{(1)}(3,2)[P^{(2)}(1,3) + P^{(2)}(1,1)P^{(1)}(2,3)] + P^{(2)}(3,1)[P^{(1)}(2,3) + P^{(1)}(2,2)P^{(2)}(1,3)]\}/P_0; \quad (36)$$

$$P^{(S4)}(3,4) = P^{(1)}(3,4) + P^{(2)}(3,4) + \{P^{(1)}(3,2)[P^{(2)}(1,4) + P^{(2)}(1,1)P^{(1)}(2,4)] + P^{(2)}(3,1)[P^{(1)}(2,4) + P^{(1)}(2,2)P^{(2)}(1,4)]\}/P_0; \quad (37)$$

$$P^{(S4)}(4,1) = P^{(1)}(4,1) + P^{(1)}(2,1)[P^{(2)}(4,1) + P^{(2)}(1,1)P^{(1)}(4,2)]/P_0; \quad (38)$$

$$P^{(S4)}(4,2) = P^{(2)}(4,2) + P^{(2)}(1,2)[P^{(1)}(4,2) + P^{(1)}(2,2)P^{(2)}(4,1)]/P_0; \quad (39)$$

$$P^{(S4)}(4,3) = P^{(1)}(4,3) + P^{(2)}(4,3) + \{P^{(1)}(4,2)[P^{(2)}(1,3) + P^{(2)}(1,1)P^{(1)}(2,3)] + P^{(2)}(4,1)[P^{(1)}(2,3) + P^{(1)}(2,2)P^{(2)}(1,3)]\}/P_0; \quad (40)$$

$$P^{(S4)}(4,4) = P^{(1)}(4,4) + P^{(2)}(4,4) + \{P^{(1)}(4,2)[P^{(2)}(1,4) + P^{(2)}(1,1)P^{(1)}(2,4)] + P^{(2)}(4,1)[P^{(1)}(2,4) + P^{(1)}(2,2)P^{(2)}(1,4)]\}/P_0, \quad (41)$$

where $P_0 = 1 - P^{(2)}(1,1)P^{(1)}(2,2)$.

The superscripts S4, 1, and 2 refer to the total P matrix, P matrix of the left-hand SAW structure, and P matrix of the right-hand SAW structure, respectively. By an SAW structure, one can mean both a single electrode and a group of electrodes with a known total P matrix. The above relationships can be used for calculating the input admittance of a TCRF with arbitrarily varying polarity of the electrodes connected to the IDT buses, varying period, apodized electrodes in the IDT

and reflecting arrays, and a realistic distribution of the surface current (charge) over the IDT's electrodes [3].

INPUT ADMITTANCE OF A TCRF

We will calculate the input admittance of a TCRF with the modified COM method. We represent either IDT as a device with two electric and six acoustic inputs (outputs). Then, the equivalent circuit of a TCRF can be represented as shown in Fig. 2. The transducers IDT-1 and IDT-2 are under potentials U_1 and U_2, which generate currents I_1 and I_2, respectively, through the devices. The complex amplitudes of acoustic waves incident on IDT-1 and IDT-2 from the left are denoted as R_n^{A1}; the corresponding reflected waves, as S_n^{A1}. The complex amplitudes of acoustic waves incident on IDT-1 and IDT-2 from the right are denoted as S_n^{A2}; the respective reflected waves, as R_n^{A2}. The complex amplitudes of acoustic waves incident on the reflecting arrays RA1-L (RA2-L) and RA1-R (RA2-R) from the left are denoted as R_n^{B1} and R_n^{C1}, respectively; those of the corresponding reflected waves, as S_n^{B1} and S_n^{C1}. The complex amplitudes of acoustic waves incident on the reflecting arrays from the right are denoted as S_n^{B2} and S_n^{C2}, respectively; those of the corresponding reflected waves, as R_n^{B2} and R_n^{C2}.

Acoustic coupling between the transducers IDT-1 and IDT-2, as well as between the reflecting arrays RA1-L, RA2-L and RA1-R, RA2-R, is defined by the transverse acoustic field distribution in the structure, which is described by expressions (11) and (12).

To obtain the components of the $[P^{(IDT)}]$ matrices, which describe the TCRF's transducers in the absence of the reflecting arrays RA1-L (RA2-L) and RA1-R (RA2-R), we write equations similar to Eqs. (22) for a kth electrode of IDT-1 and IDT-2, set $p_K = d_T$, and successively multiply the matrix components according to (25)–(41). As a result, we obtain

$$\begin{vmatrix} S^{A1}(\omega) \\ R^{A1}(\omega) \\ \Delta I_1(\omega) \\ \Delta I_2(\omega) \end{vmatrix} = \begin{vmatrix} P^{(IDT)}(1,1) & P^{(IDT)}(1,2) & P^{(IDT)}(1,3) & P^{(IDT)}(1,4) \\ P^{(IDT)}(2,1) & P^{(IDT)}(2,2) & P^{(IDT)}(2,3) & P^{(IDT)}(2,4) \\ P^{(IDT)}(3,1) & P^{(IDT)}(3,2) & P^{(IDT)}(3,3) & P^{(IDT)}(3,4) \\ P^{(IDT)}(4,1) & P^{(IDT)}(4,2) & P^{(IDT)}(4,3) & P^{(IDT)}(4,4) \end{vmatrix} \times \begin{vmatrix} R^{A2}(\omega) \\ S^{A2}(\omega) \\ U_1 \\ U_2 \end{vmatrix}. \quad (42)$$

The P matrices of the reflecting arrays RA1-L (RA2-L) and RA1-R (RA2-R) are derived from Eqs. (22) and (25)–(41) in a similar manner. It should be taken into account that the reflecting arrays do not reverse the forward- and backward-traveling acoustic waves, and electric potentials are not applied to them. Therefore, in deriving the P matrices of the reflecting arrays, we must put $\xi_K = 0$, $U_0 = 0$, and $C_2 = 0$. In this case, the matrix elements $P(1, 3)$, $P(2, 3)$, $P(3, 1)$, $P(3, 2)$, $P(3, 3)$, $P(1, 4)$, $P(2, 4)$, $P(4, 1)$, $P(4, 2)$, and $P(4, 4)$ vanish. Then, with $p_K = d_R$, for the P matrix of the reflecting array RA1-L (RA2-L), we obtain

$$\begin{vmatrix} S^{B1}(\omega) \\ R^{B2}(\omega) \end{vmatrix} = \begin{vmatrix} P^{(AL)}(1,1) & P^{(AL)}(1,2) \\ P^{(AL)}(2,2) & P^{(AL)}(2,2) \end{vmatrix} \begin{vmatrix} R^{B1}(\omega) \\ S^{B2}(\omega) \end{vmatrix}, \quad (43)$$

and for the P matrix of the reflecting array RA1-R (RA2-R),

$$\begin{vmatrix} S^{C1}(\omega) \\ R^{C2}(\omega) \end{vmatrix} = \begin{vmatrix} P^{(AR)}(1,1) & P^{(AR)}(1,2) \\ P^{(AR)}(2,2) & P^{(AR)}(2,2) \end{vmatrix} \begin{vmatrix} R^{C1}(\omega) \\ S^{C2}(\omega) \end{vmatrix}. \quad (44)$$

The matrix $[P^{(D1)}]$, which describes the gap between the reflecting array RA1-L (RA2-L) and IDT, and the matrix $[P^{(D2)}]$, which describes the gap between the IDT and reflecting array RA1-R (RA2-R), can be obtained from Eqs. (22) and (25)–(41) with $\xi_K = 0$, $U_0 = 0$, $C_2 = 0$, $r_K = 0$, and $W_K = 0$. In this case, only the elements $P(1, 2)$ and $P(2, 1)$ will be nonzero. Figure 2 shows that $R^{(A1)} = R^{(B2)}\exp(-jkd_1)$, $S^{B2} = S^{A1}\exp[jkd_1]$, $R^{C1} = R^{A1}\exp(-jkd_1)$, and $S^{A2} = S^{C1}\exp(jkd_1)$, where d_1 is the spacing between the IDT and reflecting array.

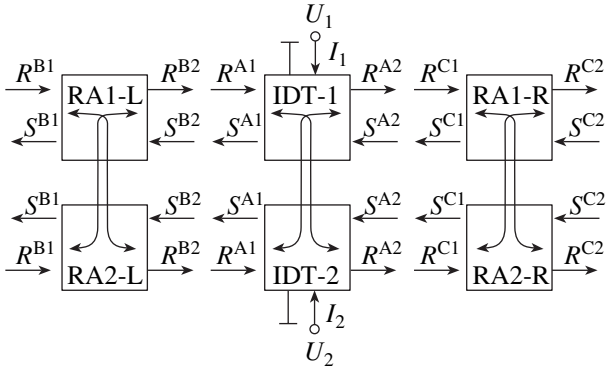


Fig. 2. Equivalent acoustoelectric circuit of a TCRF.

By successively multiplying the matrices in accordance with formulas (25)–(41),

$$[P^{(\mathrm{L})}] = [P^{(\mathrm{D1})}] \times [P^{(\mathrm{AL})}] \times [P^{(\mathrm{IDT})}],$$

$$[P^{(\mathrm{R})}] = [P^{(\mathrm{AR})}] \times [P^{(\mathrm{D2})}],$$

and

$$[P^{(\mathrm{S4})}] = [P^{(\mathrm{L})}] \times [P^{(\mathrm{R})}],$$

we obtain the components of the matrix $[P^{(\mathrm{S4})}]$, which describes the transducers in the presence of the reflecting arrays RA1-L (RA2-L) and RA1-R (RA2-R).

The quantities $P^{(\mathrm{S4})}(3, 3)$, $P^{(\mathrm{S4})}(3, 4)$, $P^{(\mathrm{S4})}(4, 3)$, and $P^{(\mathrm{S4})}(4, 4)$ have the meaning of admittances. One should keep in mind that they are obtained for an nth waveguide mode. Extracting these quantities from the $[P^{(\mathrm{S4})}]$ matrix, we come to

$$\begin{vmatrix} I_{1,n}(\omega) \\ I_{2,n}(\omega) \end{vmatrix} = \begin{vmatrix} Y_n(1,1) & Y_n(1,2) \\ Y_n(2,1) & Y_n(2,2) \end{vmatrix} \begin{vmatrix} U_1 \\ U_2 \end{vmatrix}, \quad (45)$$

$$(46)$$

where $Y_n(1, 1) = P^{(\mathrm{S4})}(3, 3)$, $Y_n(1, 2) = \pm P^{(\mathrm{S4})}(3, 4)$, $Y_n(2, 1) = P^{(\mathrm{S4})}(4, 3)$, and $Y_n(2, 2) = \pm P^{(\mathrm{S4})}(4, 4)$ are the components of the input admittance matrix of the filter due to an nth waveguide mode.

Taking into account the sign of U_2, one should use a plus sign before $Y_n(1, 2)$ and $Y_n(2, 2)$ for symmetric modes; a minus sign, for antisymmetric ones.

Since this theory considers only linear interactions (the transformation of forward- and backward-traveling SAWs and also the reflection and propagation of SAWs), the total current through the IDT is equal to the sum of the currents related to each of the modes excited plus the high-frequency current through the IDT static capacitance:

$$I_1(\omega) = \sum_n [I_{1,n}(\omega)] + j\omega(C_2/2)U_1 \sum_{k=1}^{N_\mathrm{T}} \eta_{1K}, \quad (47)$$

$$I_2(\omega) = \sum_n [I_{2,n}(\omega)] + j\omega(C_2/2)U_2 \sum_{k=1}^{N_\mathrm{T}} \eta_{2K}. \quad (48)$$

Note that the number of waves being excited and their symmetry are determined by solving the dispersion relation derived from boundary conditions (5) and (8). Then, after the input admittance matrix components due to each of the waveguide modes excited in the TCRF have been calculated from Eqs. (45)–(48), they are merely summed to find the components of the TCRF's total input admittance:

$$\begin{vmatrix} I_1(\omega) \\ I_2(\omega) \end{vmatrix} = \begin{vmatrix} Y(1,1) & Y(1,2) \\ Y(2,1) & Y(2,2) \end{vmatrix} \begin{vmatrix} U_1 \\ U_2 \end{vmatrix}, \quad (49)$$

$$(50)$$

where

$$Y(1, 1) = \sum_n P_n^{(\mathrm{S4})}(3, 3) + j\omega(C_2/2) \sum_{k=1}^{N_\mathrm{T}} \eta_{1K}; \quad (51)$$

$$Y(1, 2) = \sum_n \pm P_n^{(\mathrm{S4})}(3, 4); \quad (52)$$

$$Y(2, 1) = \sum_n P_n^{(\mathrm{S4})}(4, 3); \quad (53)$$

$$Y(2, 2) = \sum_n \pm P_n^{(\mathrm{S4})}(4, 4) + j\omega(C_2/2) \sum_{k=1}^{N_\mathrm{T}} \eta_{2K}. \quad (54)$$

Note that the plus sign in the expressions for $Y(1, 2)$ and $Y(2, 2)$ should be used for symmetric modes; a minus sign, for antisymmetric modes.

The equivalent electrical circuit of a TCRF has the form of a Π-section composed of passive elements with complex admittances. Using the equivalent electric circuit and input admittances of (51)–(54), the S parameters of a TCRF can easily be found by the standard technique.

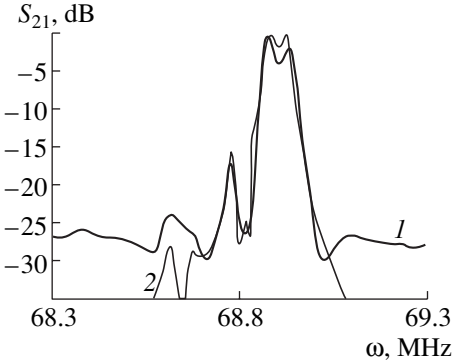


Fig. 3. Transmission coefficient of an unmatched two-mode TCRF: (*1*) experiment and (*2*) simulation.

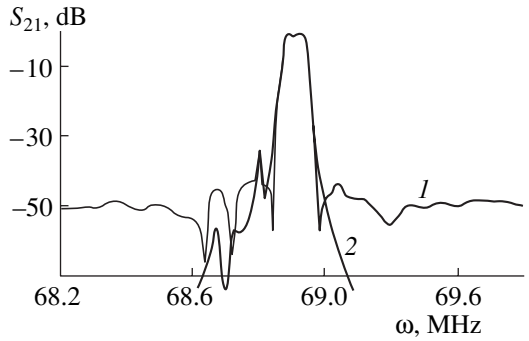


Fig. 4. Transmission coefficient of two matched TCRFs connected in series: (*1*) experiment and (*2*) simulation.

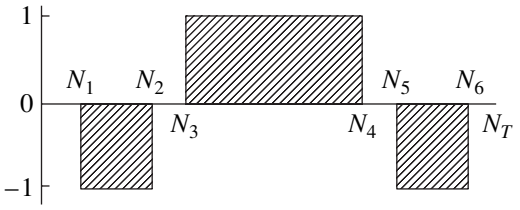


Fig. 5. Electrode polarity reversal in the four-mode TCRF.

EXPERIMENTAL RESULTS AND SIMULATION

To experimentally validate the theory developed above, we will use the results reported in [7], where a TCRF with the same polarity of electrode pairs is considered. Therefore, it can be simulated with both the conventional COM theory (see, e.g., [1]) and the theory presented in this paper. The filter was designed for a center frequency of 68.9 MHz using the conventional COM theory and the procedure of optimizing the parameters W, G, N_T, and d_T/d_R. To provide the required bandwidth (≈80 kHz), *ST, X* quartz was taken as a piezoelectric substrate material. The basic parameters of the filter's layout were $d_T = 22.835$ μm, $d_R = 22.791$ μm, the number of electrodes in the transducer IDT-1 (and IDT-2) $N_T = 561$, the number of electrodes in the reflecting array RA1 (and RA2) $N_R = 200$, $W = 6.3\lambda_0$, $G = \lambda_0$, thickness of the film A1 $0.018\lambda_0$, $d_T = 0.875\lambda_0$, and $W_K = 5\lambda_0$, where λ_0 is the acoustic wavelength at the center frequency.

The measured transmission coefficient (S_{21}) of the filter connected to a 300-Ω transmission line without matching elements is shown in Fig. 3 (curve *1*) together with the result of simulation based on the theory proposed in this paper (curve *2*). The –3-dB bandwidth and the insertion loss were measured to be 85 kHz and 5.5 dB, respectively. The plateau at –27 dB in the experimental curve is attributed to direct leakage. Note that the input (output) impedance of the filter measured with a R4-37 network analyzer is ≈300 Ω.

In order to improve the out-of-band suppression, we connected two filters (with $W = 6.3\lambda_0$ for the first filter and $5.9\lambda_0$ for the second) in series. The input and output of either filter were matched with the help of 0.6-μH inductances, and an emitter follower was connected to the input of the first filter. The frequency response of the two filters that was taken using the 50-Ω R4-37 network analyzer is shown in Fig. 4 (curve *1*) together with the result of simulation based on the theory proposed (curve *2*). The insertion loss and –3-dB bandwidth were measured to be 7 dB and 82 kHz, respectively. The out-of-band suppression in the range 10–200 MHz was no less than 43 dB except for the narrow (about 10-kHz-wide) –36-dB peak at 68.770 MHz.

To validate our theory in the case when four modes are simultaneously excited in a system of two weakly coupled resonators, we took advantage of the experimental data reported in [6] for a TCRF with four reversals of the polarity of electrodes connected to the IDT contact buses. The basic parameters of the filter's layout were as follows: the total number of electrodes in the transducer IDT-1 (IDT-2) $N_T = 999$, the number of electrodes in the reflecting array RA1 (RA2) $N_R = 250$, $W = 8\lambda_0$, $G = 2\lambda_0$, thickness of the film A1 $0.017\lambda_0$, and $d_1 = 0.875\lambda_0$. The piezoelectric substrate material was *ST, X* quartz. The polarity of electrode pairs is reversed at $N_1 = 91$, $N_2 = 139$, $N_3 = 326$, $N_4 = 674$, $N_5 = 861$, and $N_6 = 909$. Electrodes from the first to N_1th, from N_2th to N_3th, from N_4th to N_5th, and from N_6th to N_Tth were dummy. The electrode polarity variation is illustrated in Fig. 5. If we assume that the electrode polarity in the central part of the IDT is positive, i.e., +1, then the polarity of electrodes from N_1th to N_2th and from N_5th to N_6th is negative, i.e., –1.

The measured transmission coefficient of the filter connected to a 50-Ω transmission line without matching elements is shown in Fig. 6a (data from [6]). The results of simulation based on our theory are illustrated

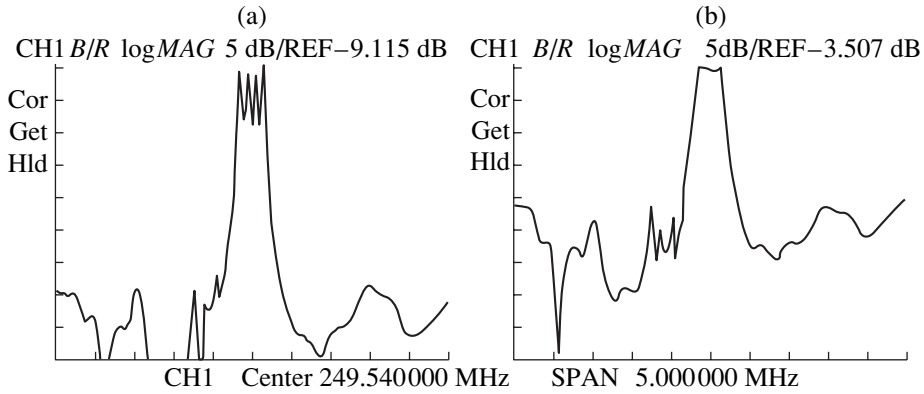


Fig. 6. Measured transmission coefficient of (a) the unmatched and (b) matched four-mode TCRF [6].

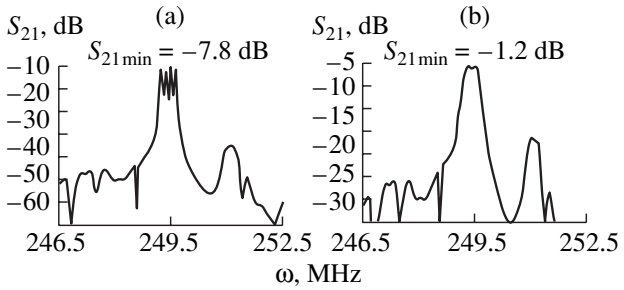


Fig. 7. Transmission coefficient of (a) the unmatched and (b) matched four-mode TCRF calculated by the method proposed in this paper. The filter layout is the same as in [6].

in Fig. 7a (theoretical results in [6] are absent). The measured transmission coefficient of the filter matched to the transmission line is plotted in Fig. 6b (data from [6]), while Fig. 7b shows the curve calculated with our theory. Good agreement between the experiment [6] and the proposed theory is observed. The theoretical insertion loss is slightly lower than the experimental values, presumably because our calculations ignored the ohmic loss in the IDT electrodes and reflecting arrays.

REFERENCES

1. S. V. Birykov, G. Martin, V. G. Polevoi, *et al.*, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 612 (1995).
2. V. F. Dmitriev, Zh. Tekh. Fiz. **72** (7), 95 (2002) [Tech. Phys. **47**, 1024 (2002)].
3. V. F. Dmitriev, Zh. Tekh. Fiz. **72** (9), 93 (2002) [Tech. Phys. **47**, 1158 (2002)].
4. V. F. Dmitriev, Zh. Tekh. Fiz. **72** (11), 83 (2002) [Tech. Phys. **47**, 1427 (2002)].
5. G. Martin, in *Proceedings of IEEE Ultrasonics Symposium* (1999), p. 15.
6. G. Martin and B. Waill, in *Proceedings of IEEE Ultrasonics Symposium* (1997), p. 37.
7. V. F. Dmitriev and N. P. Osipova, in *Proceedings of International Forum on Wave Electronics and Its Application* (St. Petersburg, 2000), p. 360.

Translated by A. Khzmalyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 239–244. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 107–112.

ELECTRON AND ION BEAMS, ACCELERATORS

Diagnostics of a Hollow Relativistic Electron Beam from the Transfer Function of a Converting Target

N. G. Mordasov

Federal State Unitary Enterprise Research Institute of Devices, Lytkarino, Moscow oblast, 140080 Russia

e-mail: risi@tsr.ru

Received February 27, 2002

Abstract—Methods for measuring angular, geometrical, and energy characteristics of a hollow relativistic electron beam using the transfer function of a converting target are developed. A relationship between the anisotropy parameters of bremsstrahlung behind the target, angular characteristics of electrons, and geometrical parameters of the electron beam in the target plane is established. The energy balance equation that describes the interaction between the electron beam and converting target is studied. The transfer function of the target and its behavior versus electron energy are found. The time and integral characteristics of the electron beam are examined. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

15-MeV electron accelerators are widely used for studying nuclear-radiation-related effects and in applications concerned with the generation of high-intensity pulsed photon radiation [1, 2]. The radiation is generated in a target that converts a part of the electron energy to bremsstrahlung (converting target). The parameters of the bremsstrahlung depend directly on the type and thickness of the target, as well as on the energy, current, and angle of incidence of electrons.

Although electron beams in accelerators have been used for a long time, a reliable correlation between the bremsstrahlung characteristics and electron beam parameters, which could be used to optimize the electron beam conversion and obtain given bremsstrahlung fields, is still lacking. According to [3], the energy yield I of bremsstrahlung per electron (without considering self-absorption in a target with an atomic number Z and assuming that normally incident electrons of energy E are totally absorbed) is given by

$$I = ZE^2/(ZE + 6.53 \times 10^3), \text{ MeV}. \quad (1)$$

The strong dependence between the energy yield of bremsstrahlung and the electron energy has stimulated the development of methods for finding the energy and current characteristics of electron beams by measuring the bremsstrahlung characteristics. In particular, it was proposed [4] to measure the time distributions of the electron energy $E(t)$ and current $J(t)$ with a set of Compton detectors that have essentially different energy dependences of the sensitivity to photon radiation.

A set of dose-rate detectors with replaceable filters was used in [5], where electrons were assumed to normally strike the target. The disadvantages of the techniques suggested in [4, 5] are the inadequacy of calibration and measurement conditions and a narrow energy range available because of fundamental limitations. In [6], a diode collimating array was used to study the time evolution of the electron beam radial structure by tracing the bremsstrahlung yield. The feasibility of evaluating the parameters of a continuous cylindrical electron beam from those of secondary radiation was carefully investigated in [7]. These studies have made it possible to work out an algorithm for the synthesis of the target transfer functions to estimate the energy, current, and angle of incidence of electrons in the energy range below 3 MeV. The associated technique uses a set of at least four dose-rate detectors with identical energy dependences of the sensitivity (which are placed at fixed distances from the target at different angles to the axis of electron beam transport (G = 0–180°)) and electron beam current meters.

The measuring methods developed are indirect and assume that bremsstrahlung measurements are exact. Under certain conditions, this approach is apparently justified.

At present, there are no data for using similar methods to measure the characteristics of high-current hollow relativistic electron beams. In this paper, we fill this gap to a certain extent, suggesting experimental approaches to reconstructing geometrical, angular, and energy characteristics of hollow electron beams in the plane of a converting target.

1063-7842/03/4802-0239$24.00

SUBJECT OF INVESTIGATION AND EXPERIMENTAL TECHNIQUE

The problem of increasing the bremsstrahlung yield under laboratory conditions was partially solved by using linear pulsed accelerators (of LIU type) with a hollow electron beam [8]. Hollow electron beams are accelerated and transported in axisymmetric electric and magnetic fields. The limiting current $J_{\lim}$ that can be transported under these conditions is given by

$$J_{\lim} = J_0(\gamma^{2/3} - 1)^{3/2}[1 + 2\ln(l/r)], \quad (2)$$

where $J_0 = 17$ kA is the Alfvén current; γ is the relative electron energy; and l and r are the thickness and effective radius of the electron beam, respectively.

The arrangement of the accelerator's inductors built on radial lines provides a wide range of electron energy. The electron energy spread $\Delta\gamma$ over the beam cross section depends on the beam current J as

$$\Delta\gamma/\gamma = 0.5lJ/rJ_{\lim}. \quad (3)$$

For the LIU-10 accelerator, the maximum electron energy is 12 MeV, the beam current is ≤40 kA at a cathode diameter of 8 to 12 cm, and the pulse width is $\leq 3 \times 10^{-8}$ s.

When a high-current hollow electron beam is transported through an accelerating tube in an axisymmetric magnetic field, the particle velocity vector makes a certain angle with the magnetic induction vector. This angle depends on the self-consistent structure of the self-electromagnetic fields of the beam and external fields. The complex motion of an electron in such fields can be represented as the motion of a point (guiding center) in appropriate coordinates plus the rotational motion about this center. The path of the electron's guiding center is a helix with a radius depending on the guiding center azimuth velocity, i.e., on the self-electric field (radial component) and resultant magnetic field (axial component), while the path of an individual electron is a helix with a Larmor radius that is wound on the guiding center path.

Such a representation of the particle motion is not rigorous but visualizes its influence on the bremsstrahlung field behind the target. Relative to the beam transport axis, the electron motion has the axial, $V_\parallel$; azimuthal, V_a; and radial, V_r, velocity components, which are related to the angle of incidence Θ as

$$\Theta = \arctan\{\sqrt{(V_a^2 + V_r^2)/V_\parallel}\} = \arctan\sqrt{(\tan^2\beta + \tan^2\varphi)}. \quad (4)$$

With reproducible axisymmetric magnetic and accelerating electric fields, the angle of incidence depends primarily on its energy and, for a several-megaelectronvolt energy range, varies within a radiation pulse only slightly; i.e., under these conditions, an electron beam during a pulse can be characterized by the effective angle of incidence Q.

The method for determining the angular and geometric characteristics of a hollow relativistic electron beam incident on a converting target relies on the fact that the energy and angular distributions of bremsstrahlung leaving an absorbent with an atomic number Z depends on its thickness D, electron energy E, and effective angle of incidence Q. The basic property of the angular distribution of the bremsstrahlung generated when a relativistic electron is deflected in the field of a nucleus or bound electron is that the photons travel predominantly in the direction of the primary electron, mostly within a cone whose apex angle is inversely proportional to the relative electron energy γ [2]. The angular distribution of bremsstrahlung generated in converting targets of a thickness comparable to or greater than the primary electron mean free path is affected by multiple scattering and depends on the electron energy loss. An analytical expression describing the bremsstrahlung from such absorbents has not been derived. Based on qualitative considerations, one can expect that, as the target thickness increases, so does the angular distribution width and the angle of maximum bremsstrahlung yield may not coincide with Q when $Q \neq 0$. The difference Y between Q and the angle of maximum bremsstrahlung yield increases with Z, D, and Q and decreases with increasing E. Under particular measurement conditions and geometry, Y can be obtained from independent experiments with monoenergetic electron beams. Velocity components responsible for the bremsstrahlung exit angle (as in (4)) in the case of a hollow electron beam are determined from the positions of maxima in the bremsstrahlung dose rate distribution over j ($j \geq 3$) planes at different distances from the target that are normal to the electron beam transport axis [9] (for example, with the help of dose-rate detectors).

In the coordinate planes, the sequences of the bremsstrahlung dose rate distribution maxima produced within a pulse are represented by normal (relative to the electron beam transport axis) sections of a one-sheet hyperboloid (at $Q \neq 0$). The geometric parameters of these sections are used to find the coefficients a, c, and h of the equation of one-sheet hyperboloid (in the canonical form) whose rectilinear generatrices are collinear with the direction of the maximum bremsstrahlung yield. Therefore, the components of the effective (over a pulse) angle of electron incidence given by (4) can be expressed through the coefficients a, c, and h of the canonical equation of one-sheet hyperboloid as follows:

$$\beta = \arctan(a/c) + Y_\beta, \quad (5)$$

$$\varphi = \arctan\{ah/c\sqrt{(c^2 + h^2)}\} + Y_\varphi. \quad (6)$$

Using this technique, one can also determine the one-pulse effective radius r of the cross section of a hollow electron beam in the target plane:

$$r = (a/c)\sqrt{(c^2 + h^2)}. \quad (7)$$

To calculate the energy characteristics of the beam's electrons, consider the energy balance equation for the interaction of the beam with a homogeneous converting target whose thickness meets the total absorption condition for primary electrons. For the electron energies of interest, the beam energy W can be represented by the sum of the energy absorbed in the target (W_{ab}), that carried away by bremsstrahlung (W_{br}), backscattered electron radiation (W_Q), δ-electron radiation (W_δ), and photoneutron radiation (W_{ph}); i.e.,

$$W - W_{ab} - W_Q - W_\delta - W_{br} - W_{ph} = 0, \quad (8)$$

The energy W of the electron beam per radiation pulse can be expressed by the time-dependent beam current $J(t)$ and electron energy $E(t)$:

$$W = e^{-1}\int_0^\tau J(t)E(t)dt, \quad (9)$$

where e is the electron charge.

According to [7, 10, 11], the energy $E(t)$ of electrons incident on the target at the time instant t within the width τ of an electron radiation pulse can be represented through the dynamic characteristic of the target transfer function ($[U(t) = P(t)/J(t)]$ as follows:

$$E(t) = b^{-1}[P(t)/J(t)]^{-n(t)}\big|_{Q = \text{const}}. \quad (10)$$

Here, $P(t)$ is the time-dependent bremsstrahlung dose rate at a particular point of the field and

$$n(t) = \alpha_0 + \beta_0[P(t)/J(t)] + \gamma_0[P(t)/J(t)]^2 + \ldots, \quad (11)$$

where b, α_0, β_0, and γ_0 are the coefficients depending on measurement conditions and geometry.

The parameters W_{ab}, $J(t)$, $P(t)$, and Q are determined experimentally during a radiation pulse. The components W_{ph}, W_{br}, W_Q, and W_δ from (8) can be expressed as a function of W given by (9) with relationship (10) taken into account.

Consider in detail the components of Eq. (8) for a single-layer converting target. Due to their small mass, electrons are intensely scattered in a material; therefore, the probability that they will reflect from the target is fairly high, especially at $Q \neq 0$. By analogy with (9), the component of Eq. (8) that refers to the beam energy loss W_Q during a pulse due to electron backscattering can be represented as

$$W_Q = e^{-1}\int_0^\tau R(Q, E, Z)J(t)E_{av}(t)dt, \quad (12)$$

where $R(Q, E, Z)$ is the total electron backscattering factor and $E_{av}(t)$ is the average energy of backscattered electrons.

The total electron backscattering (reflection) factor R depends largely on their energy E, effective angle of incidence Q on the target, and atomic number Z of the target material and can be analytically represented as [12]

$$R = V + \{1 - V\}\{1 - (1 + \sigma)^{-0.5} \times [(1 - 1.73\cos Q)/(1 + 1.73 \times \{\cos Q(1 - \sigma)/(1 + \sigma)\}^{-0.5})]\}\exp(-1/\sigma), \quad (13)$$

where V is the tabulated function of the electron scattering angle [12] and $\sigma \approx (Z + 1)/E$ is the number of events of electron scattering by angles on the order of 1.

The average energy E_{av} can be estimated as [12]

$$E_{av} \approx 0.6E[1 - E(1 - \sin Q)/(Z + 1)]. \quad (14)$$

The component W_{br} in Eq. (8), which represents the electron beam energy carried away from the target by bremsstrahlung, is determined from the radiation yield by introducing the efficiency coefficient for target bremsstrahlung [2] (which depends on the atomic number of the target and electron energy) and by estimating the self-absorption of the bremsstrahlung radiation in the target; i.e.,

$$W_{br} = e^{-1}KZ\int_0^\tau W(t)E(t) \times \exp[-\mu\{E(t)\}(D - d\{E(t)\})/\cos Q]dt, \quad (15)$$

where K is a coefficient in MeV^{-1}; $\mu\{E(t)\}$ is the mass coefficient of photon absorption in the converting target that is averaged over the photon energy spectrum; and D and $d\{E(t)\}$ are, respectively, the thickness of the target and the depth where the effective generation of photons starts [2].

The component W_δ in Eq. (8), which represents the energy taken away from the target by δ electrons, is estimated in accordance with [13]. Here, it is taken into account that the bremsstrahlung spectrum, excluding the characteristic radiation, is almost independent of the target's atomic number Z and the relative bremsstrahlung spectrum, of the electron energy. Thus, W_δ depends on the bremsstrahlung cross section alone and is no greater than 0.2% of W in (9) for target materials with $Z \leq 74$.

The analysis of the photoneutron yield [2] shows that W_{ph} in (8) is no greater than 0.1% of the electron beam energy W for the vast majority of materials; i.e., W_δ and W_{ph} in Eq. (8) can be viewed as corrections. As follows from the aforesaid, in the electron energy range considered, the significant components of Eq. (8) are W, W_{ab}, W_{br}, and W_Q, of which W_{br} and W_Q can be expressed analytically through the components of the

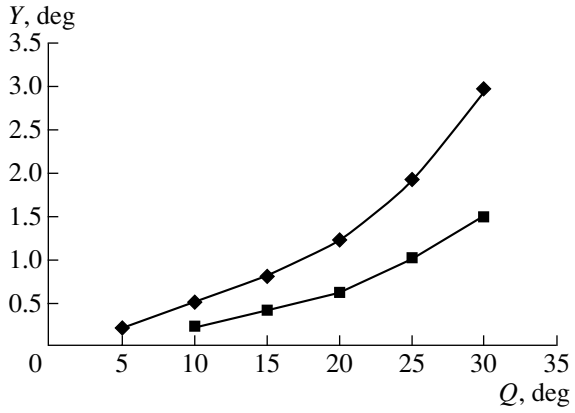


Fig. 1. Y versus Q at E = (■) 4 and (◆) 12 MeV.

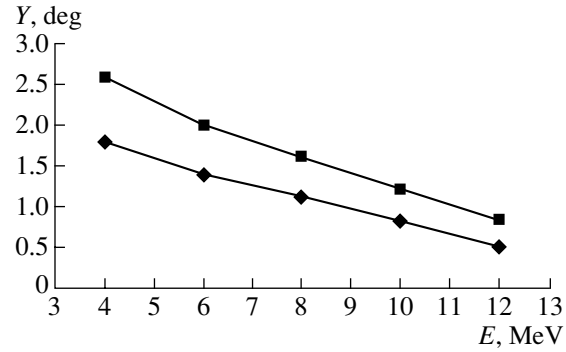


Fig. 2. Y versus E at Q = (■) 15° and (◆) 30°.

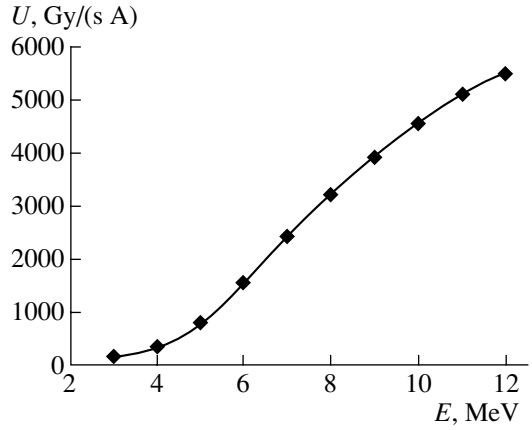


Fig. 3. Transfer function $U(E)$ of the converting target.

parameter W given by (9) and W_{ab} is determined experimentally. The errors in W_{br} and W_Q calculated and in the parameters $P(t)$, $J(t)$, Q, and W_{ab} measured may significantly affect the solution. The ratios of W_Q, W_{br}, and W_{ab} to W depend on the atomic number of the target. The lower Z, the smaller W_Q/W and W_{br}/W and the greater W_{ab}/W; i.e., the significance of W_Q and W_{br} in energy balance equation (8) can be minimized by appropriately choosing Z of the target. To find the coefficients b, α_0, β_0, and γ_0 from (10) and (11), we will use the method stated in [14]. In this method, $P(t)$, $J(t)$, Q, and W_{ab} are first measured by applying a series of $d \geq 4$ radiation pulses from the accelerator that differ significantly in maximum electron energy, with the experimental geometry and conditions remaining the same, and then system of nonlinear equations (8) is solved.

EXPERIMENTAL

The above techniques were used to study the angular and energy characteristics of electrons in a hollow beam at the exit of an LIU-10 linac in order to optimize its operating conditions for obtaining a high-intensity uniform bremsstrahlung field. The converting target was made of iron (Z = 26 and D = 7.2 g cm^{-2}), and the bremsstrahlung dose rate was measured with an SD2-01 scintillation detector made of P-15 plastic and placed 1 m behind the target on the electron beam axis. The beam current was measured with a sectional Rogowski loop. The current and dose-rate measuring channels were synchronized and had identical time resolutions. The beam energy absorbed in the target, W_{ab}, was determined with a copper thermal resistor, which was regularly applied on the surface of the target and had a uniform surface resistance per unit length [14]. The coordinates of the dose distribution maxima were found with IS-7 thermoluminescent glass detectors and then processed with a computer program. In independent experiments with continuous cylindrical beams of monoenergetic electrons with the same experiment geometry, we determined the parameter Y versus energy and angle of electron incidence over given ranges of these parameters (Figs. 1, 2) with the help of thermoluminescent detectors.

With an induction of an axisymmetric magnetic field applied to the accelerator channel B = 0.4 T and a maximum electron energy from 8 to 12 MeV, the effective angle of incidence changed from 16° (β = 14° and φ = –8°) to 13° (β = 11° and φ = –7°). The negative value of φ means that the electron beam radially diverges near the target. The value $\varphi \neq 0$ characterizes the deviation of the electron beam from the cylindrical form.

As follows from the experiment, the induction B = 0.4 T of the external axisymmetric magnetic field, which prevents the spread of the electron beam in the accelerating tube (φ = 0), decreases substantially near the target because of the edge and skin effects, which changes the angular characteristics of the electrons. When the external magnetic field near the target is increased to 0.55 T, the component of the effective angle of incidence $\varphi \approx 0$ and the component β increases to 18° (for E = 12 MeV).

The dynamic transfer function $U(t)$ of the converting target was constructed by machine processing the measured distributions $P(t)$ and $J(t)$. From the experi-

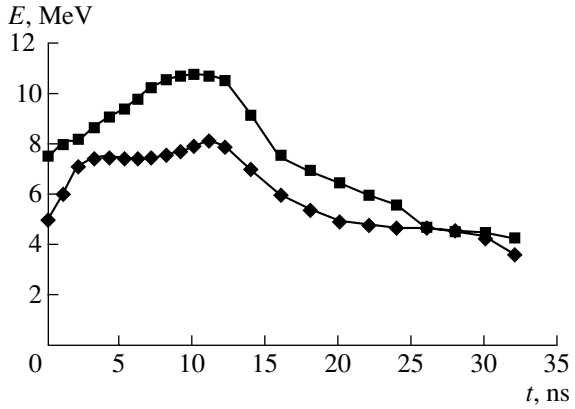


Fig. 4. Time distribution of the electron energy in a radiation pulse: (■) intermediate acceleration mode and (◆) acceleration mode optimized in electron energy.

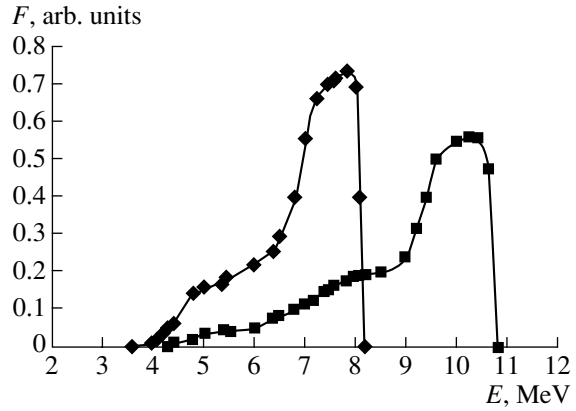


Fig. 5. Electron energy spectra in a radiation pulse. The acceleration modes are the same as in Fig. 4.

mental results and in view of the measurement errors involved in P, J, Q, and W_{ab}, we obtained a solution to Eq. (8) when the number of equations in the system is $d \geq m$ (m is the number of unknowns b, α_0, β_0, γ_0, …):

$$b = 0.047, \quad \alpha_0 = 7.23 \times 10^{-2},$$

$$\beta_0 = 0.91 \times 10^{-2}, \quad \gamma_0 = -2 \times 10^{-4}.$$

These results allow us to find the transfer function of the target $U = P/J$ versus the electron energy, $U(E)$, shown in Fig. 3. When $U(E)$ and dynamic transfer function $U(t)$ are known, it is possible to calculate $E(t)$, $W(t)$, and the electron energy spectrum $F(E)$ in the target plane. The functions $E(t)$ and $F(E)$ are plotted in Figs. 4 and 5, respectively, for two energy regimes. The discrepancy in the rise times of the electron energy pulse when a different number of the inductors is used (Fig. 4) and the tail observed in the low-energy region of the integral electron energy spectra (Fig. 5) make it possible to estimate the temporal mismatch between the accelerating fields of the inductors.

CONCLUSION

Thus, we established a correlation between the parameters of the bremsstrahlung field and those of a hollow electron beam incident on a converting target whose thickness meets the total reflection condition for primary electrons.

It is shown that, when the bremsstrahlung is produced by a hollow electron beam transported in an axisymmetric magnetic field, the components of the angle of incidence of relativistic electrons on the target can be determined from the geometrical parameters of the maxima of the bremsstrahlung dose rate distribution in several planes. The effect of the angle of incidence and energy of electrons on this correlation is clarified. The transfer function of the target is defined as the ratio of the bremsstrahlung dose rate at a point on the electron beam axis to the beam current. The solution of Eq. (8) is, in essence, the calibration of the measuring system, primary current converters, bremsstrahlung dose rate, and target in a given geometry. This calibration specifies the transfer function of the target versus the electron beam energy, $U(E)$. The procedure employs the absolute calibration method measuring the energy W_{ab} released in the target and may be omitted in subsequent measurements.

The techniques developed in this paper are used to study the geometrical, angular, and energy characteristics of a hollow electron beam transported to a converting target in an axisymmetric magnetic field. The time behavior of the electron energy and current allow one to analyze the acceleration process, improve the synchronization of the accelerating field inductors, and thus optimize the angular and geometrical characteristics of the electron beam in the target plane. This information is sufficient to calculate the bremsstrahlung energy spectrum for any time interval (within the pulse width) and can be used to find ways of acting on a hollow electron beam that provide a bremsstrahlung field with specific desired properties (dose, dose rate, irradiation uniformity, equidistant planes of equal dose, etc.).

REFERENCES

1. A. N. Didenko, V. P. Grigor'ev, and Yu. P. Usov, *Powerful Electron Beams and Their Application* (Atomizdat, Moscow, 1977).
2. V. P. Kovalev, *Secondary Radiations in Electron Accelerators* (Atomizdat, Moscow, 1979).
3. V. F. Baranov, *Electron Radiation Dosimetry* (Atomizdat, Moscow, 1974).
4. J. V. Walker and J. Stevens, Rev. Sci. Instrum. **45** (2), 16 (1974).

5. M. I. Ivanov, V. M. Kazakov, O. V. Kozlov, *et al.*, At. Energ. **45**, 280 (1978).
6. S. V. Marichev and V. D. Sazhin, *Diagnostics of Charged Particle Beams in Accelerators* (Radiotekh. Inst. Akad. Nauk SSSR, Moscow, 1984), pp. 12–23.
7. N. I. Zavada, P. P. Komarov, V. A. Tsukerman, and E. N. Donskoĭ, *Physics and Technology of Pulsed Ionizing Radiation Sources to Study Fast Processes*, Ed. by N. G. Makeev (VNIIÉF, Sarov, 1996), pp. 56–95.
8. A. I. Pavlovskiĭ, V. S. Bosamykin, *et al.*, Dokl. Akad. Nauk SSSR **250**, 1117 (1980) [Sov. Phys. Dokl. **25**, 120 (1980)].
9. N. G. Mordasov and N. I. Muratov, USSR Inventor's Certificate No. 1,276,101. G01T1/29 (1985).
10. N. G. Mordasov and N. I. Muratov, USSR Inventor's Certificate No. 1,410,679 (1981).
11. N. G. Mordasov, *Radiation Hardness of Electron Systems: Hardness 2001* (Paims, Moscow, 2001).
12. M. I. Ryazanov and I. S. Tilinin, *Surface Investigation by Particle Backscattering* (Énergoatomizdat, Moscow, 1985).
13. A. F. Akkerman, M. Ya. Grudskiĭ, and V. V. Smirnov, *Gamma-Induced Secondary Electron Emission from Solids* (Énergoatomizdat, Moscow, 1986).
14. N. G. Mordasov, USSR Inventor's Certificate No. 1,737,994 (1992).

Translated by A. Khzmalyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 245–249. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 113–117.

ELECTRON AND ION BEAMS, ACCELERATORS

Generation of Electron Beams in a Multicathode Secondary-Emission Source

N. I. Aĭzatskiĭ, V. N. Boriskin, A. N. Dovbnya, V. V. Zakutin, V. A. Kushnir, V. A. Mitrochenko, N. G. Reshetnyak, V. P. Romas'ko, Yu. Ya. Volkolupov, and M. A. Krasnogolovets

Uskoritel' Research Complex, Kharkov Institute of Physics and Technology, National Science Center, Kharkov, 61108 Ukraine

Received April 22, 2002

Abstract—A study is made of the generation of electron beams in a system consisting of eight secondary-emission cathodes arranged at regular intervals in the azimuthal direction inside a coaxial cylindrical anode in crossed electric and magnetic fields. In this system with an azimuthally nonuniform electric field, secondary-emission multiplication of electrons is realized and beam generation is achieved. With a cathode voltage of ~37 kV and a magnetic field of ~3000 Oe, the total current of all the beams amounts to ~35 A, the microperveance of each beam being ~0.7 μA/V$^{3/2}$. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In recent years, studies of electron sources with cold cathodes operating in the regime of secondary emission in crossed electric and magnetic electric fields have aroused considerable interest [1–6]. The principle of beam generation in secondary-emission systems in crossed fields is well known. When the voltage drops, primary electrons moving in a decreasing electric field and constant magnetic field along cycloidal trajectories gain energy and bombard the cathode. If the electric field at the cathode decreases very rapidly, the electrons gain enough energy to knock out secondary electrons (with a secondary yield σ above unity). At this instant, the number of both bombarding and knocked-on electrons increases in an avalanche-like manner. The number of electrons near the cathode increases until it reaches a certain critical value, at which the dynamic equilibrium is established, the secondary-electron multiplication becomes steady state, and beam generation begins. Primary electrons in the system can be produced by field emission, emission from dielectric impregnations on the cathode surface, or emission from an auxiliary hot cathode [1, 2].

For generating an electron flow by secondary-emission multiplication, the crossed electric and magnetic fields must be properly produced near the cathode. In view of the boundary conditions near the metal surface, the electric field is always normal to the cathode surface. Therefore, in any system of cylindrical secondary-emission cathodes (including those possessing no azimuthal symmetry throughout the entire interelectrode gap), conditions exist such that an electron beam may be generated by secondary-emission multiplication.

A system consisting of rod anodes and secondary-emission cathodes was studied in detail in [7]. In this system, the rod anodes are positioned at the intersection of the axes forming squares in the *xy* plane, and the rod cathodes are located between them at the intersections of the diagonals and form a set of cells. The azimuthal profile of the electric field in each cell varies from almost uniform (near the surfaces of the rod cathodes) to strongly nonuniform (at a distance from the cathode equal to or longer than half the distance from the rod anodes). Such a design allows the regular arrangement of the cells with secondary-emission cathodes and, consequently, of the electron beams, whose number can be increased either in one axial direction (to form a linear beam) or in both directions (to form an arbitrary beam configuration in the *xy* plane). With four cells, a cathode voltage of ~28 kV, and a magnetic field of ~2200 Oe, electron beams with a total beam current of ~22 A were produced.

It is of interest to study a nonuniform system similar to the linear system with rod anodes and cathodes, but in cylindrical geometry. Such a system consists of a coaxial cylindrical anode and secondary-emission cathodes arranged at regular intervals in the azimuthal direction between the anode cylinders. In this paper, we study the generation of electron beams in a multicathode secondary-emission system with a coaxial anode in crossed electric and magnetic fields.

EXPERIMENTAL SETUP AND MEASUREMENT TECHNIQUE

A schematic of the experimental device is shown in Fig. 1. Specially shaped voltage pulses with an amplitude of up to 100 kV, a flat-top duration of ~5 μs, and a repetition rate of 12–15 Hz were fed from pulse modulator *1* (where *FL* is the forming line, *L* is the charging

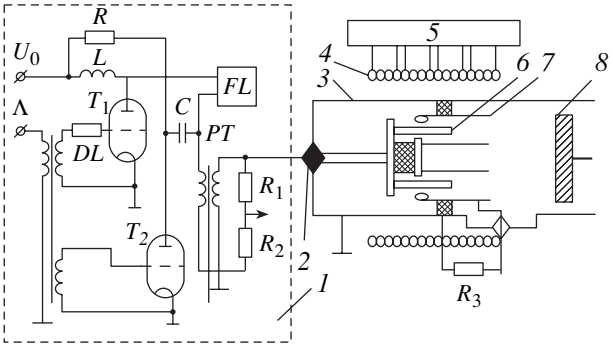


Fig. 1. Schematic of the experimental device.

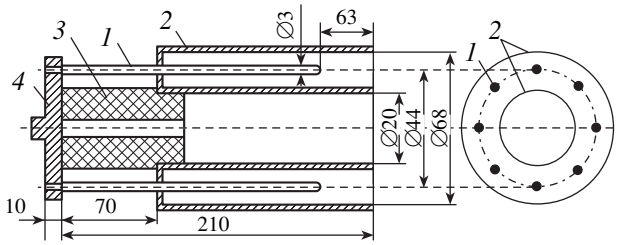


Fig. 2. Multicathode system: (*1*) secondary-emission cathodes, (*2*) anode, (*3*) insulator, and (*4*) metal disk.

choke coil, U_0 is the charging voltage, R is the charging resistor, *DL* is the delay line, C is the forming capacitor, and *PT* is the step-up pulse transformer) to cathodes *6* of the system through lead-in insulator *2*. Anode *7* was grounded through resistor R_3. The pulse top necessary for the beam generation was formed at the cathodes by adding two pulses: a short pulse, produced by a separate generator with thyratron T_2, and a flat-top voltage pulse, generated by a modulator with thyratron T_1 [8]. The duration of the voltage drop was ~0.3 μs, and the drop steepness was ~150 kV/μs. We used a nine-cell forming line with a wave impedance of 12 Ω and a full width at half maximum of ~7 μs.

The multicathode system was placed in vacuum chamber *3*, which was evacuated to $\sim 10^{-6}$ torr. The magnetic field necessary for the generation and transportation of the beam was created by solenoid *4*, consisting of four sections and powered by dc source *5*. The strength and longitudinal profile of the magnetic field in the vacuum chamber could be controlled by varying the currents in the solenoid sections. The inhomogeneity of the longitudinal magnetic field in the system and in the transport channel (up to the Faraday cup) was ±8%.

We measured the cathode voltage, the electron beam current at the Faraday cup, and the anode current. The parameters of the voltage pulse were measured with the help of resistive divider R_1R_2, and the anode current was measured with the help of measuring resistor R_3. The Faraday cup *8* was a segment of a coaxial line with an impedance of ~12 Ω. The line length was ~400 mm, and the cylinder diameter was 120 mm. The stainless-steel end surface of the Faraday cup, which was exposed to the beam, was cooled with water. The beam current in the measurement circuit reached its steady-state value in 1.0–1.5 ns. The beam dimensions were measured from beam prints on an aluminum or a molybdenum foil. In order to estimate the reproducibility of the shape and amplitude of the cathode voltage pulses, as well as the beam current at the Faraday cup, in addition to oscilloscope measurements, a double-channel computer-based system was also used.

The design of the system with a cylindrical anode, in which secondary-emission cathodes are arranged, is shown in Fig. 2. It can be seen that the system contains eight copper cathodes *1* (5 mm in diameter) mounted on metal disk *4*. Coaxial anode *2* is mounted on the same disk through insulator *3*. The outside diameter of the anode cylinder is 68 mm, and the inside diameter is 20 mm. The stainless-steel anode cylinders are jointed with a metal flange, which has holes to insert the cathodes.

EXPERIMENTAL RESULTS AND DISCUSSION

In the experiments, we achieved secondary-emission electron multiplication and the generation of electron beams. The total beam at the outlet from the gun consisted of 8 electron beams. At a voltage amplitude of ~37 kV and magnetic field of 3000 Oe, the total beam current amounted to ~35 A and the anode current attained 1.5–2 A. Stable beam generation was obtained with a voltage drop steepness of ~150 kV/μs.

Figure 3a shows typical oscillograms of the cathode voltage and the total beam current at the Faraday cup. It can be seen that the beam generation begins and the leading edge of the beam-current pulse forms when the cathode voltage drops. The measured rise time of the beam-current pulse is 10–12 ns.

Figure 3b shows oscillograms of the anode current and the total beam current at the Faraday cup. It can be seen from Fig. 3b that the anode current is similar in shape to the beam current and amounts to ~3% of the beam current. In the oscillogram of the anode current, we can see an initial signal related to the differentiation of the spike of the voltage pulse. We also measured the dependence of the beam current at the Faraday cup on the cathode voltage. This dependence obeys the 3/2 law. In these measurements, the magnetic field corresponded to the maximum amplitude of the total beam current.

In experiments, it was found that the design of the cathode insertion into the anode significantly affected the generation stability and the beam amplitude. To study this effect, we examined the beam generation in two versions of the cathode insertion. In the first version, the flange joining the anode cylinders was posi-

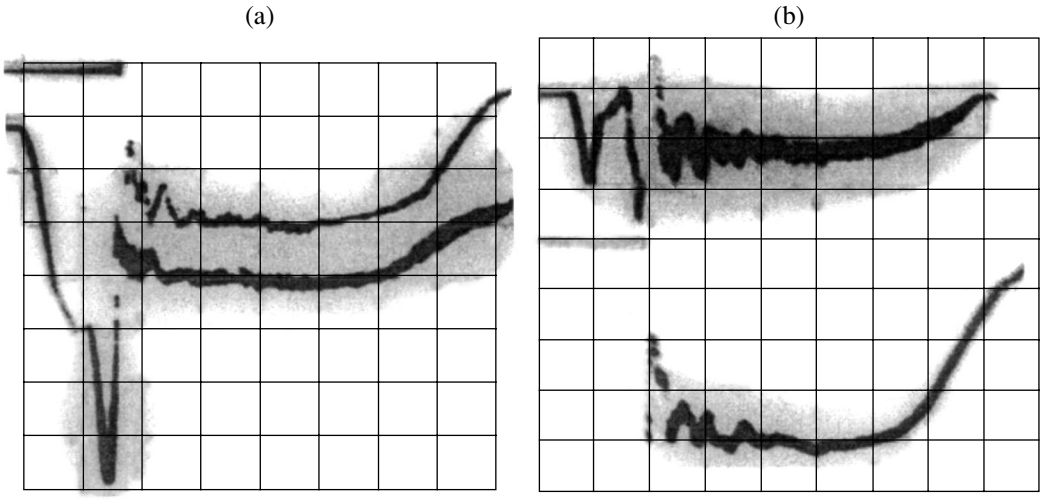


Fig. 3. Typical oscillograms of (a) the voltage pulse at the cathode (bottom) and the current pulse at the Faraday cup (top) (the vertical scales are 8 kV/division and 8 A/division, respectively; the horizontal scale is 1 μs/division) and (b) beam-current pulse at the Faraday cup (bottom) and the gun anode current (top) (the vertical scale is 8 A/division for the beam current and 1 A/div for the anode current; the horizontal scale is 1 μs/division). The magnetic field is ~2500 Oe.

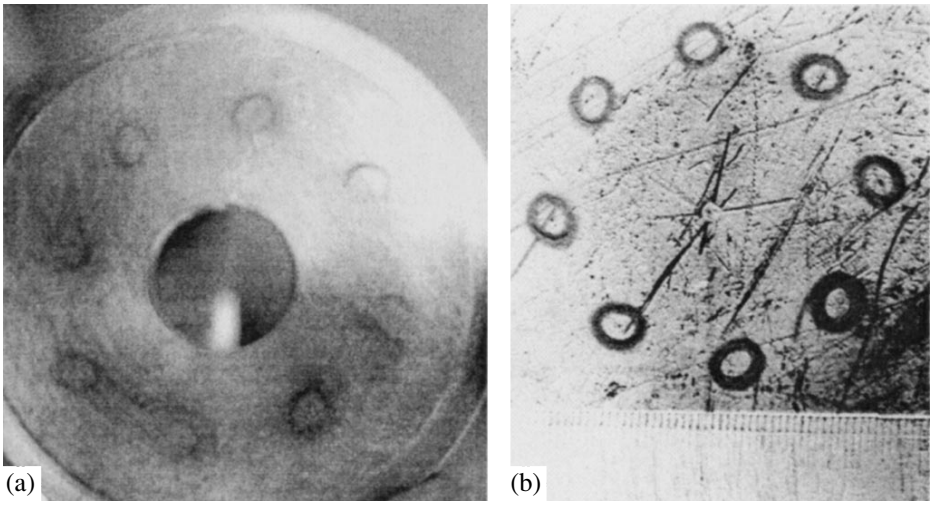


Fig. 4. Beam prints (a) at the outlet from the anode and (b) at a distance of 72 mm from the anode.

tioned on the cathode-insertion side. In the second version, the joining flange was positioned on the beam-exit side, whereas the cathodes were inserted into the anode through the gap between the inner and outer cylinders, and the anode had a conical part at the cathode-insertion site (in order to increase the electric strength). The experiments showed that the first version ensured the stable generation of the beam current. In the second version, generation was also stable, but the beam current was lower. In this version, the beam-current pulse consisted of two parts: an initial flat-top part (corresponding to stable generation) with a duration of 0.5–1.5 μs and a subsequent (unstable) part with current spike amplitudes of ~5 to 10 A. As the magnetic field increased, the duration of the flat-top part increased (to 1.5 μs), but the spiky character of the subsequent beam generation remained the same.

The above effects are explained by the fact that, changing the design of the cathode insertion, we change secondary-emission conditions in the initial stage of the beam formation. The change in the strength and distribution of the edge electric field in the sites where the cathodes are inserted into the anode results in a change in both the number of primary electrons and the conditions under which these electrons gain energy. With a high electric field, the number of electrons is larger, they gain energy more rapidly, and their final energy is higher. With a lower electric field, both the number of the primary electrons and their energy decrease; as a result, the beam generation becomes

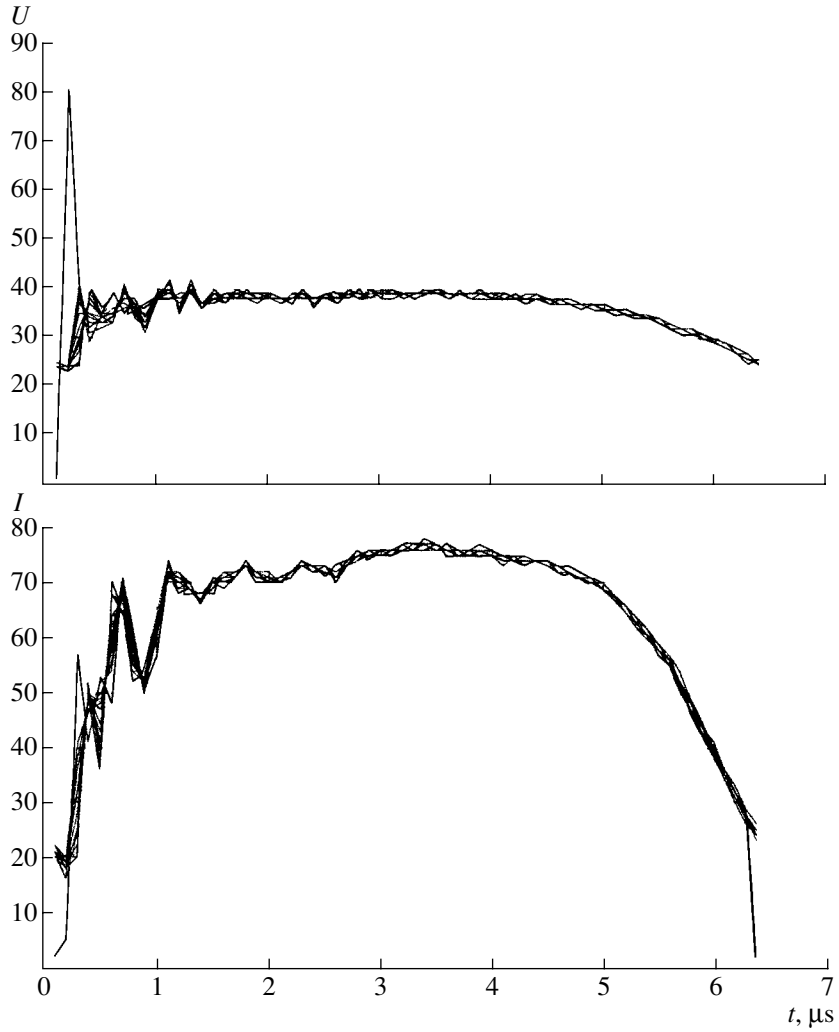


Fig. 5. Waveforms of the cathode voltage pulses and the beam current at the Faraday cap.

unstable. On the other hand, the multicathode system is a system with a nonuniform field. Calculations show that, in this case, the character of electron motion changes substantially and self-sustained secondary-emission multiplication can even become impossible. Hence, the cylindrical part of the cathode insertion into the anode acts as a stabilizing element that determines the development of the secondary-emission processes. These results show that the secondary-emission processes can be initiated by introducing a section in which the electric field is strongly nonuniform.

The transverse dimensions and positions of the electron beams were measured from their prints on targets placed at different distances from the anode (Fig. 4). These measurements showed that, at the outlet from the anode, the system forms eight elliptical electron beams with a major axis of 6.5–7.0 mm and minor axis of 5.5–6.0 mm. The beam positions correspond to the cathode positions. The beam-print thickness is ~1 mm. The ellipse axes are inclined at an angle of ~45° to the circle on which the cathodes are arranged (Fig. 1). When the magnetic-field direction is reversed, the inclination of the ellipse axes changes by 90°. The ellipse dimensions are the same both at the exit from the anode unit and at a distance of 72 mm from the exit.

A physical reason why we obtain such prints is as follows. The electric-field distribution over the transverse cross section of the system is nonuniform. Calculations show that the radial electric field near the cathode surface is nonuniform within ~20% in the azimuthal direction: the shorter the distance to the anode, the higher the electric field. This causes radial electron drift and deforms the shape of the beam cross section, which transforms from a circle into an ellipse inclined at a certain angle about the beam axis.

We note that the magnetic field in the multicathode system is higher than the magnetic field of a single magnetron gun with nearly the same dimensions (with a cathode diameter of 5 mm, an anode diameter of 26 mm, a cathode voltage amplitude of 37 kV, and a

magnetic field of 1900–2000 Oe) by a factor of 1.5. For instance, stable beam generation in the multicathode system with a cathode voltage amplitude of 36 kV was achieved at a magnetic field of 3000 Oe.

This might be the result of a change in the electric-field distribution in the system under study. In the absence of an electron cloud, the spatial region where the electric field near the cathode is uniform in the transverse direction is small. When an electron sheath (with a substantial space charge) is formed, the electric field near the cathode is forced out from the sheath and the size of the region where the electric field is uniform decreases. As a result, a higher magnetic field is required for beam formation.

The cathode voltage pulses and the beam current at the Faraday cup were also measured with a double-channel computer-based system. Experimental points were obtained by sampling from running oscillograms with a sampling period of 200 ns (at a voltage amplitude of 34 kV and beam current of 30 A). This allowed us to estimate the beam-generation stability in 16 successive pulses, i.e., the amplitude stability from pulse to pulse. These results are presented in Fig. 5, which demonstrates a rather stable beam generation.

CONCLUSION

It has been demonstrated that crossed-field electron guns with rod cathodes and a nonuniform electric field can be used to achieve self-sustained electron-beam generation due to secondary electron emission. The electric-field distribution near the sites where the rod cathodes are inserted into the anode system considerably affects the characteristics of the generated beam. The electron flow from the gun makes up N distinct beams, where N is the number of cathodes (in our case $N = 8$). At the outlet from the anode unit, the beam cross section is elliptical in shape, and this shape changes only slightly during beam propagation.

REFERENCES

1. Yu. Ya. Volkolupov, A. N. Dovbnya, V. V. Zakutin, *et al.*, Zh. Tekh. Fiz. **71** (2), 98 (2001) [Tech. Phys. **46**, 227 (2001)].
2. Y. M. Saveliev, W. Sibbett, and D. M. Parkes, Phys. Plasmas **4**, 2319 (1997).
3. Yu. Ya. Volkolupov, A. N. Dovbnya, V. V. Zakutin, *et al.*, Zh. Tekh. Fiz. **71** (3), 78 (2001) [Tech. Phys. **46**, 348 (2001)].
4. A. V. Agafonov, V. P. Tarakanov, and V. M. Fedorov, Vopr. At. Nauki Tekh., Ser.: Yad.-Fiz. Issled., Nos. 2–3, 134 (1997).
5. A. V. Agafonov, V. P. Tarakanov, and V. M. Fedorov, Vopr. At. Nauki Tekh., Ser.: Yad.-Fiz. Issled., No. 4, 11 (1999).
6. Yu. Ya. Volkolupov, A. N. Dovbnya, V. V. Zakutin, *et al.*, Zh. Tekh. Fiz. **71** (9), 134 (2001) [Tech. Phys. **46**, 1196 (2001)].
7. N. I. Aĭzatskiĭ, Yu. Ya. Volkolupov, A. N. Dovbnya, *et al.*, Pis'ma Zh. Tekh. Fiz. **27** (23), 25 (2001) [Tech. Phys. Lett. **27**, 991 (2001)].
8. A. Dovbnya, N. Reshetnyak, V. Romasko, *et al.*, in *Proceedings of Particle Acceleration Conference (PAC-2001) Chikago, 2001*, Vol. 5, p. 3759.

Translated by N. Larionova

Technical Physics, Vol. 48, No. 2, 2003, pp. 250–256. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 118–124.

SURFACES,
ELECTRON AND ION EMISSION

Variety of LaSrMnO Structures Induced by Growth Conditions and Laser Irradiation

Z. A. Samoĭlenko*, V. D. Okunev*, E. I. Pushenko*, T. A. D'yachenko*, A. Cherenkov*, P. Gierlowski, S. J. Lewandowski**, A. Abal'oshev**, A. Klimov**, and A. Szewczyk****

** Galkin Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 83114 Ukraine*
e-mail: okunev@host.dipt.donetsk.ua
*** Instytut Fizyki PAN, Al. Lotnikov 32/46, 02-668 Warszawa*
Received February 20, 2002

Abstract—Crystallographic phase transitions in perovskite-like LaSrMnO metallic oxides are studied. The transitions are induced when internal stresses generated during film synthesis (at temperatures between 450 and 730°C) vary (decrease or increase) upon subsequent irradiation by a KrF laser emitting in the UV range. As the synthesis temperature T_s grows, the rhombohedral-to-orthorhombic phase transition occurs at 650–670°C. The resistivity is shown to be either temperature-independent, $\rho(T) = \text{const}$, at $T < T_{crit}$, or varies and reaches a maximum, $\rho(T) = \rho_{max}$, at the Curie temperature T_c. Optical transmission spectra taken at photon energies $\hbar\omega = 0.5–2.5$ eV exhibit both a high (0.8–0.9) and low (0.1–0.3) transmission coefficient t, depending on the synthesis temperature. As follows from X-ray diffraction data, the laser irradiation causes a phase transition only in LaSrMnO films grown at $T_s < 650$°C. Phases of different size scales appear: the long-range-order orthorhombic matrix and mesoscopic-range-order rhombohedral clusters are observed in the films grown at $T_s = 450–550$°C and the rhombohedral matrix with orthorhombic clusters, in the films grown at $T_s = 550–650$°C. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It has been reported [1–5] that the unit cell of lanthanum manganese oxides like LaSrMnO (which attract researchers' attention because of the giant magnetoresistance effect) may vary from cubic to rhombohedral.

The variability of the unit cell structure in perovskite-like metallic oxides is explained by the fact that growth-related internal stresses may be accumulated and accordingly released in a different way in the multicomponent system, causing the crystallographic self-organization of one or another order.

Taking into account [6, 7] and our own experience, the structure evolution in order of increasing lattice distortions is generally the following: cubic ⟶ tetragonal ⟶ orthorhombic ⟶ rhombohedral, or

$$a = b = c, \quad a \neq b = c,$$
$$\alpha = \beta = \gamma = 90°, \quad \alpha = \beta = \gamma = 90°,$$
$$a \neq b \neq c, \quad a = b = c,$$
$$\alpha = \beta = \gamma = 90°, \quad \alpha = \beta = 90°, \quad \gamma \neq 90°.$$

MATERIAL AND INVESTIGATION TECHNIQUES

Samples were obtained by the sputtering of a $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ target with a pulsed KrF laser (τ = 25 ns, energy density $\Phi = 3.0$ J/cm^2, oxygen pressure in the working chamber $P = 300$ mtorr). The films were deposited on $SrLaGaO_4$, $Nd_3Ga_5O_{12}$, and $Gd_3Ga_5O_{12}$ substrates in the temperature range 450–730°C and subsequently irradiated by the same laser at lower energy densities ($0.1 < \Phi < 0.15$ J/cm^2; the number of pulses was varied from 5 to 500).

The film structure was examined by the X-ray diffraction method using CrK_α radiation combined with the photometry method [8], which makes it easy to detect X rays diffusely scattered by clustered solid solutions (i.e., our objects; Fig. 1). Electrical measurements were carried out in the temperature interval 4.2–300 K by standard techniques. Transmission spectra were taken at room temperature in the photon energy range $\hbar\omega = 0.5–5.0$ eV with an SP 700C spectrophotometer.

RESULTS AND DISCUSSION

Film synthesis temperature. The stressed state of the film structure was varied by varying the LaSrMnO synthesis temperature T_s between 450 and 730°C. The lower limit is close to the amorphous–crystalline (disorder–order) phase transition (T = 350–400°C); the upper, to the experimentally found optimum temperature ($T \approx 750$°C) for the synthesis of good films (with a Curie temperature $T_c > 300$ K).

Substrates for epitaxial growth. Substrates used in the experiment had a much lesser effect on the film stressed state than the synthesis temperature. We used

$Gd_3Ga_5O_{12}$, $SrLaGaO_4$, and $Nd_3Ga_5O_{12}$ perovskite-like single-crystal substrates with different interplanar spacings.

Varying these two parameters (the synthesis temperature and lattice–substrate lattice mismatch), we provided conditions for growing elastically stressed $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ layers and thus created a variety of crystalline phase structures via the self-organization of the elastically stressed multicomponent charge-inhomogeneous system. Moreover, the system is subjected to a time–temperature regime that favors diffusion processes ($450 < T_s < 730$°C, 10 min).

The rhombohedral-to-orthorhombic phase transition at $650 \le T_s \le 670$°C. Diffraction patterns taken from the $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films synthesized at T_s = 450–730°C suggest that, first, the films are crystalline (as follows from Laue reflections observed in the patterns for all diffraction angles) and, second, that, as T_s grows, the crystallographic phase states of the films vary as

$$R(3c) \xrightarrow[T_s < 650°C]{} R(3c) + O(Pnma) \xrightarrow[T_s > 670°C]{} O(Pnma);$$

in other words, the single-phase rhombohedral structure $R(3c)$ passes to the single-phase orthorhombic phase $O(Pnma)$ through the two-phase structure with comparable amounts of the phases. The integral diffraction maximum intensities for the corresponding structure groups are shown in Fig. 2. The curve $I = f(T)$ fluctuates, which indicates the labile rearranging low-coherence atomic structure of the $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ two-phase (rhombohedral + orthorhombic) state at T_s = 650–670°C (Fig. 2, curves *2* init).

Phase state in the LaSrMnO films after laser irradiation. The labile atomic order in the material under study shows up after the UV ($\hbar\omega \approx 5$ eV) laser irradiation. In the films synthesized at high temperatures ($T_s > 650$°C), the structure does not change qualitatively. However, in the films grown at lower temperatures ($T_s \le 600$°C), which are in a more stressed state, the laser irradiation splits the temperature domain of existence of the single-phase rhombohedral structure (450–600°C) into two two-phase domains with a demarcation line near T_s = 550°C (Fig. 2; curves *3*, *4*). Below 550°C, the structure consists of the orthorhombic matrix and rhombohedral clusters (basic reflections from the *O* phase and diffuse maxima from the *R* phase; Fig. 1c). Conversely, above 550°C, the matrix is rhombohedral, while the clusters are orthorhombic.

The stressed state of the LaSrMnO atomic structure revealed from diffraction patterns. The disturbance of the atomic order in the crystalline structure (appearing as ion displacements from "zero" lattice sites) is accompanied by the enhancement of the intensity I_{incoh} of X-ray incoherent scattering. This follows from Fig. 3, which shows the I_{incoh} maxima near T_s = 550 and 670°C. This pattern agrees with findings in

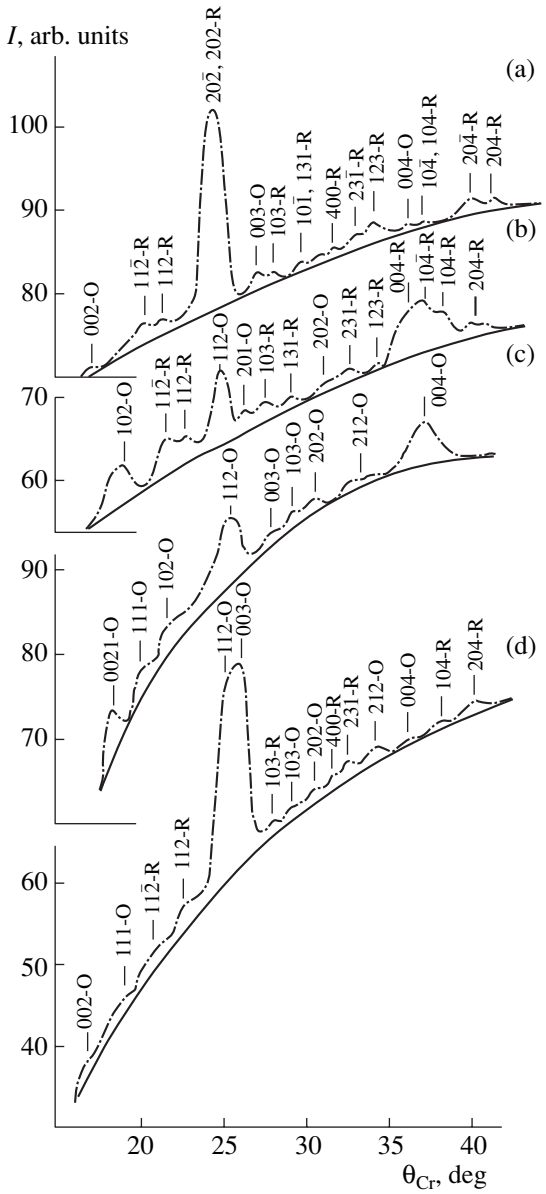


Fig. 1. Typical diffraction patterns from $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films synthesized at different temperatures T_s. (a) 450–600°C (rhombohedral phase $R3c$), (b) 650–670°C (rhombohedral $R3c$ + orthorhombic $Pn3m$ two-phase region), (c) 700–730°C (orthorhombic phase $Pn3m$), and (d) 450–550°C with irradiation (phase $Pn3m$ with $R3c$ clusters).

[9, 10] and indicates that the system becomes disordered before it acquires a new stable atomic order.

As for "near-zero" ion displacements, one should bear in mind that the resulting directions of the displacements are nonzero on average. They remain significant, reflecting the presence of either tensile or compression stresses in the material, which alternate as the synthesis temperature T_s grows. This follows from the oscillating T_s dependences of the interplanar spacing d (Fig. 4). The dependences were calculated (by the

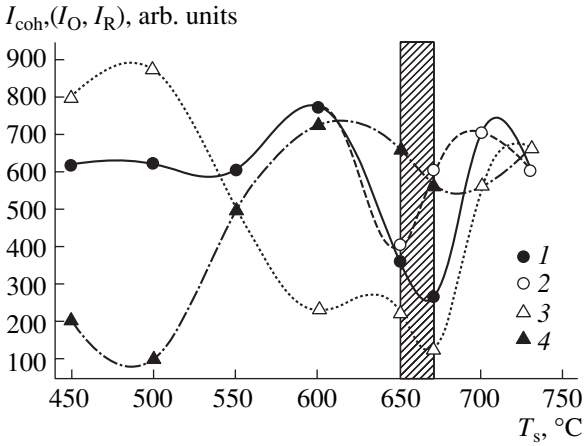


Fig. 2. Integral intensities of X-ray coherent scattering from the $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films grown at different T_s for (*2, 3*) orthorhombic and (*1*, *4*) rhombohedral phases. Curves *3* and *4* are taken after the laser irradiation.

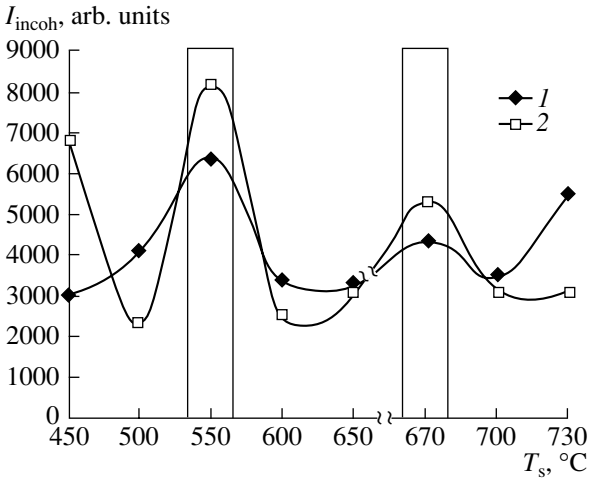


Fig. 3. Integral intensity of X-ray noncoherent scattering vs. synthesis temperature T_s for the (*1*) as-grown and (*2*) irradiated $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films.

Bragg formula $2d\sin\theta = n\lambda$) for the principal maximum from diffraction angles θ found experimentally. The discrepancy between the experimental and tabulated values of d for the $Gd_3Ga_5O_{12}$ substrate (curve *1*) in the *R* range ($T_s < 650$°C) varies between +0.08 and –0.04 Å (or +2.94 and –1.47%) relative to $d(202)$ = 2.72 Å. In the *O* range ($T_s > 650$°C), the value of d approaches 2.80 Å, which corresponds to {020} planes of the orthorhombic phase. Note that this group of *O* planes differs from the {412} planes identified at $T_s \le 650$°C (Fig. 1, curve *1*). On the crystallographic order scale, a distortion of 2% is very large. For example, studies [11] on the formation of new phase nuclei showed that if the radii of neighboring atoms differ by less than 2%, the nuclei are spherical, and when the difference exceeds 2%, the nuclei grow flat; that is, the shape of growing (isotropic or anisotropic) phase grains changes qualitatively. In our samples, as T_s rises, the rhombohedral-to-orthorhombic phase transition takes place, which agrees with the hierarchical order observed in the evolution of the perovskite-like structure: from more distorted to less distorted, i.e., from lower symmetry *R*3*c* to higher symmetry *Pnma.*

The effect of laser irradiation on the dependence $d = f(T_s)$. As follows from Fig. 4, the laser irradiation does not change the periodic run of the dependence $d(T_s)$ but inverts it relative to the dependence for the as-grown sample (cf. curves *1*, *2*). This means that the laser energy shifts the periodic variation of elastic stresses in the LaSrMnO structure with T_s, leaving it basically the same. That is, laser irradiation may favor (accelerate) structure modifications without radically changing the material properties. This is confirmed by Fig. 3, which demonstrates the phase states in the as-grown and irradiated material.

The effect of the substrates on the dependence $d = f(T_s)$. Comparing curves *1*, *3*, and *4* in Fig. 4, we see that the $SrLaGaO_4$, $Nd_3Ga_5O_{12}$, and $Gd_3Ga_5O_{12}$ substrates have both a similar and different effect on the dependence $d = f(T_s)$. The similar effect is that the curves oscillate synchronously (in phase) at $T_s < 600$°C, i.e., in the range where the rhombohedral phase of LaSrMnO forms. For the films synthesized at $T_s >$ 600°C, the difference between the curves becomes noticeable and the curve for the $SrLaGaO_4$ substrate runs in antiphase to the others.

At high synthesis temperatures ($T_s > 600$°C), the curves tend to saturate (the sinusoidal variation changes to tangential). This change is of principal character, indicating qualitative modifications in the LaSrMnO atomic and electron configurations; namely, it means that in the structurally unstable lattice (sinusoidal run of the curves $d = f(T_s)$), large stable clusters embedded in the matrix arise. These clusters generate long-range forces that are proportional to [9]

$$\mathbf{r}^{-m} \cdot \cos(2\mathbf{k}_F \cdot \mathbf{r} + \varphi),$$

where m = 1–3, φ is the phase, and the vector $2\mathbf{k}_F$ connects Fermi surface areas with the normals parallel to the radius vector **r**.

The formation of waves reduces the electron energy of the system, and its heterogeneous state becomes stable. In our material, the heterogeneous state implies the presence of mesoscopic-size (≤100 Å) clusters bounded by {00*l*} planes, because oxygen anions interact mostly with one of the metals in LaSrMnO (diffuse maxima {00*l*} in Fig. 1c).

Among {*hkl*} clusters, of most interest are those reflecting Mn–O interactions in the perovskite lattice, which are responsible for the LaSrMnO conducting and magnetic properties. According to the experimental

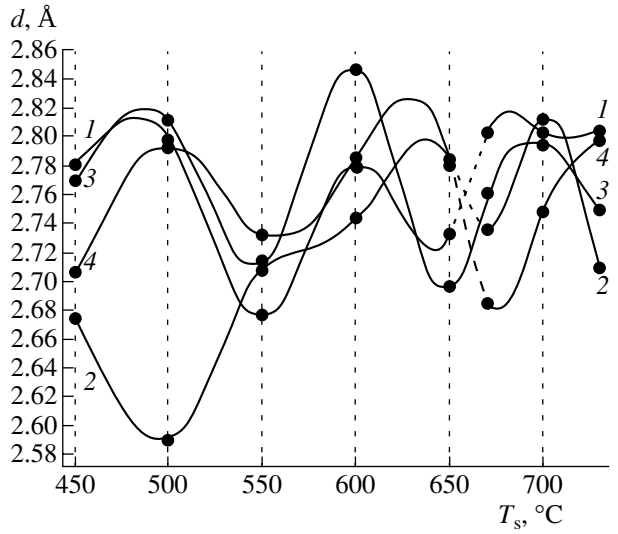


Fig. 4. Interplanar spacings d corresponding to the principal diffraction maximum vs. synthesis temperature T_s for the $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films on (*1*, *2*) $Gd_3Ga_5O_{12}$, (*3*) $Nd_3Ga_5O_{12}$, and (*4*) $SrLaGaO_4$ substrates. Curve *2* is obtained after irradiation by the KrF laser.

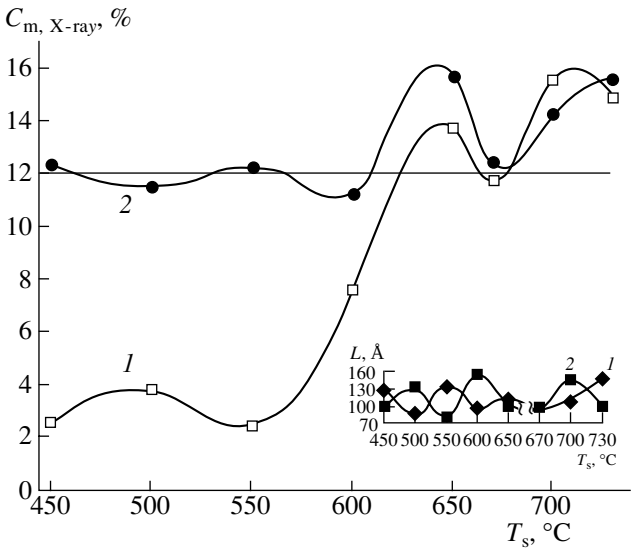


Fig. 5. Concentration C_m of the metallic phase determined from the diffraction patterns from the $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films on $Gd_3Ga_5O_{12}$ at different temperatures T_s (*1*) before and (*2*) after laser irradiation. The inset shows the sizes of the metallic clusters (*1*) before and (*2*) after the irradiation.

values of d, these are $(400)_R$ clusters in the rhombohedral phase and $(202)_O$ and $(004)_O$ clusters in the orthorhombic phase. Let us trace their temperature behavior. Having measured the integral intensities I_{coh} of all the diffraction reflections and separated out the abovementioned diffuse maxima I_m of the metallic phase, we estimated its concentration $C_m = I_m/I_{coh}$ with an accuracy of $\pm 0.05C_m$ (Fig. 5, curve *1*). It is noteworthy that $C_m < 12\%$ (at $T_s < 600$°C) and $C_m > 12\%$ (at $T_s \geq 650$°C). This difference reflects the qualitative change in the physical properties that are sensitive to the concentration of metallic granules in the insulating matrix (inhomogeneity of this type is assumed to be responsible for the conducting and magnetic properties of materials like those considered in this paper [1]). The value $C_m \approx 12\%$ is near the percolation threshold [9, 12], when charge carriers may travel along a continuous channel through metallic regions in an inhomogeneous medium and the internal energy of the system is reduced.

For the samples grown at $T_s = 650$ and 670°C, the experimental concentration of the clusters is below 12%; namely, $C_m \approx 10.8$ and 8.6%, respectively. At the same time, the metallic cluster sizes calculated from the widths of the diffuse maxima ($(400)_R$, $(202)_O$, and $(004)_O$) by the Kitaigorodsky technique vary insignificantly (as in [8]), fluctuating about 100 Å (see inset to Fig. 5). This means that the concentration C_m of metallic clusters, rather than their size, governs the variation of the LaSrMnO properties with T_s. Such a conclusion is in line with the concepts of the percolation theory [12].

Optical transmission spectra. Two types of spectra are observed (Fig. 6). For the films synthesized at $T_s <$ 600°C, which have a rhombohedral structure, the transparency in the long-wavelength ($\hbar\omega < 2$ eV) range is high. In this range, two closely spaced maxima with a minimum at $\hbar\omega = 0.9$–1.1 eV in between are observed. The regions associated with absorption by free carriers at $\hbar\omega < 0.8$ eV indicate the presence of the metallic clusters. Their concentration C_m is conveniently estimated from the relationship

$$1/t \cdot dt/d(\hbar\omega) \sim C_m p,$$

where p is the hole concentration [13].

In the films synthesized at $T_s < 600$°C, the cluster concentration thus estimated equals 2–5% [14] in accordance with the experimental data in Fig. 5.

The samples synthesized at $T_s > 600$°C are less transparent. Unlike the former samples, for which transmission spectra are well known [15], they have a minimum near $\hbar\omega = 1.5$ eV and a maximum whose position varies between 1.7 and 2.2 eV from sample to sample. According to the concepts concerning the origin of optical spectra from manganites [15], the variation of the transmission t with $\hbar\omega$ for $\hbar\omega < 1.5$ eV and >1.5 eV is related to intraband and interband transitions, respectively. For $\hbar\omega > 2.2$ eV, the density of states involved in absorption significantly grows and the transmission drops rapidly with increasing $\hbar\omega$.

The effect of the substrate on the spectra is observed only in the critical temperature range. The inset to Fig. 6 shows the spectra taken from the films grown on different substrates at $T_s = 600$°C. It is seen that the

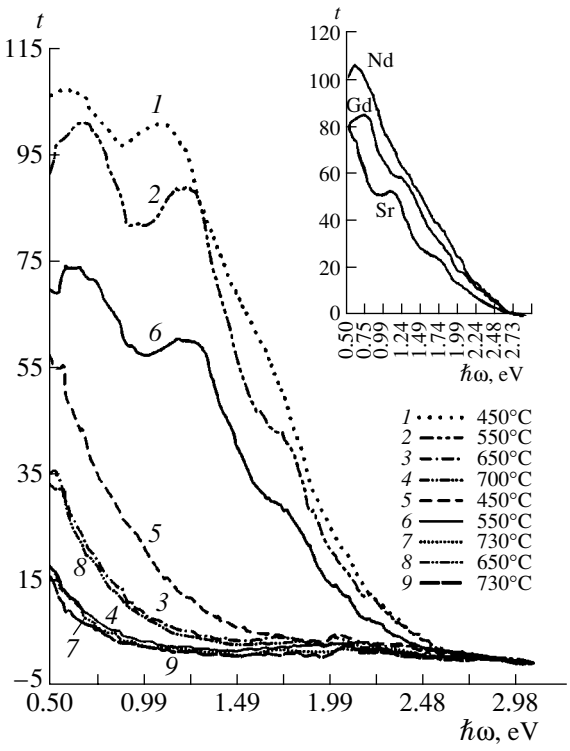


Fig. 6. Typical optical transmission spectra for the $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films synthesized at different T_s. (*1–4, 7*) Before and (*5, 6, 8, 9*) after the irradiation.

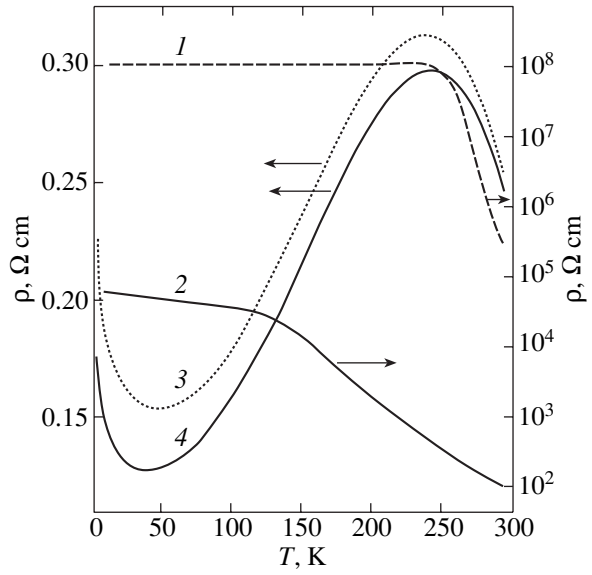


Fig. 7. Temperature dependences of the resistivity of the $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films synthesized at different T_s. (*1*) $\rho(T)$ = const at $T < T_{crit}$ (the film grown at T_s = 550°C on the $Gd_3Ga_5O_{12}$ substrate), (*2*) the same as (*1*) after the laser irradiation (30 pulses with an energy density of 0.15 J/cm^2), (*3*) $\rho(T)$ has a peak near the Curie temperature (the film grown at T_s = 730°C on the $SrLaGaO_4$ substrate), and (*4*) the same as (*3*) measured in a magnetic field of 3.8 kG.

spectrum from the film grown on $Nd_3Ga_5O_{12}$ can obviously be assigned to the spectra of the first type, while the spectrum from the film on $SrLaGaO_4$ can be assigned to the spectra of the second type. This agrees with the data in Fig. 4.

After the laser irradiation, changes in the spectra follow those in the structure: the rhombohedral phase (low T_s) loses transparency, becoming orthorhombic, while the spectra from the orthorhombic phase change insignificantly because of its stability.

Electrical conductivity. The qualitative modifications observed in the LaSrMnO films as the synthesis temperature T_s grows also affect the temperature dependence of the resistivity $\rho(T)$. For the samples grown at $T_s \le 600$°C, the dependence $\rho(T)$ is typical of semiconductors or insulators in the temperature range between T_{crit} = 250 K and room temperature (Fig. 7, curve *1*). Below T_{crit}, the curve exhibits a plateau with $\rho(T)$ = const. The laser irradiation, which causes the phase transition, reduces the resistance by three or four orders (Fig. 7, curve *2*). Also, for T between 4.2 and 100 K (where the resistivity of the as-grown samples is constant), the resistivity of the samples irradiated grows with decreasing temperature approximately by a factor of 1.5.

Plateaus in the curves $\rho(T)$ for manganites of various composition have been observed earlier [5, 14, 16–22]. The origin of these portions has been discussed in [5, 14, 16–18]. In [5], this effect is related to magnetic ordering in the films below the Néel temperature. This effect was also observed in YBaCuO, where small clusters of metallic conductivity are embedded in an insulating matrix [8, 14, 19]. With this in mind, it has been inferred that the plateaus are due to elastic electron tunneling between the clusters, which turn into quantum dots with a discrete energy spectrum at low temperatures. This is indirectly corroborated in [16], where the curves $\rho(T)$ taken from high-ρ insulating films with nanometer ferromagnetic inclusions exhibited portions with $\rho(T)$ = const.

For the films obtained at high temperatures ($T_s \ge$ 650°C), the curve $\rho(T)$ has a maximum ρ_{max} near the Curie temperature (Fig. 7, curve *3*). This maximum was also observed in [20, 21], where the "semiconductor" run ($d\rho/dT < 0$) changes to the "metallic" run ($d\rho/dT >$ 0). In the latter case, however, the similarity is formal. In fact, in our films, the range where the resistivity peaks lies between 0.2 and 1.5 Ω cm. The associated values of the conductivity σ vary between 5 and 0.7 Ω^{-1} cm^{-1}, which is much lower than the minimal metallic conductivity, which is determined by the formula [22]

$$\sigma_{min} = \frac{\pi e^2}{4z\hbar a}\left[\frac{B}{V_0}\right]^2_{crit},$$

z is the coordination number, a is the spacing between impurity centers, V_0 is the random potential amplitude, and B is the band width.

For $La_{0.6}Sr_{0.2}Mn_{1.2}O_3$ films, the value of σ_{min} must be no less than 100 Ω^{-1} cm^{-1}, which agrees with estimates made in [23]. Since the conductivity of such systems is related to d orbitals, its value is expected to approach 1000 Ω^{-1} cm^{-1} [22]. Even if one takes into account that the samples are heavily inhomogeneous and that the insulator–metal transition is of percolation character, the mean possible conductivity near the percolation threshold is one or two orders of magnitude lower than the value of σ_{min} expected. The most plausible explanation for the dependence $\rho(T)$ observed in the low-temperature range appears to be the hypothesis that, in a highly inhomogeneous medium, the hole edge of the mobility shifts relative to the Fermi level under magnetic ordering conditions [24].

The drastic decrease in the resistivity in the magnetically ordered state is associated with a decrease in the energy of activation (the material remains in the dielectric state). An external magnetic field facilitates the motion of charge carriers in the magnetic excitation "environment" [24]. Under such conditions, the resistivity of the samples in a magnetic field must decrease, which is the case in the experiments (Fig. 7, curve *4*).

To conclude, we will discuss the correlation between the electrical properties and structure of the films. As follows from the X-ray diffraction data, in the samples synthesized at T_s = 650 and 670°C, C_s is less than 12% (between 8.6 and 10.8%), i.e., below the percolation threshold. In spite of this, for T_s < 230 K, the dependences $\rho(T)$ exhibit a pronounced metallic behavior near the maximum of ρ (ρ = 1.0–1.5 Ω cm). These values are at least one or two orders of magnitude higher than those corresponding to the minimum of the metallic conductivity ($\rho = 10^{-3}$–10^{-2} Ω cm). According to today's concepts, a decrease in the LaSrMnO resistivity is due to magnetic ordering in the material. Since this takes place at C_m < 12%, it is obvious that magnetic ordering occurs at metallic (ferromagnetic) phase concentrations below the percolation threshold. The magnetically ordered state is the result of the enhanced interaction between crystallographically similar clusters (bounded by {004} planes in our case). The metallic behavior of the dependence $\rho(T)$ is observed when a continuous channel through metallic regions between the electrodes is still absent and the electrical conductivity of the films is limited by carrier tunneling between isolated but already numerous regions of metallic conductivity. Insulating spacers make a major contribution to the resistivity.

CONCLUSIONS

Our results show that phase transitions take place in local areas (clusters) of LaSrMnO solid solution, which may show themselves as quantum dots at low temperatures. Schematically, the phase transitions can be described as follows.

(1) As the synthesis temperature grows ($T_s \geq$ 650°C), elastic stresses and diffusion initiate the $R \longrightarrow O$ phase transition.

(2) Strains cause electronic density redistribution and the local (in the clusters) insulator–metal ($T_s <$ 600°C, the basic phase $R(3c)$ phase transition. At low temperatures, the clusters turn into a set of tunnel-coupled quantum dots with $\rho(T)$ = const.

(3) The reconfiguration of the lattice and electronic subsystems results in the paramagnet–ferromagnet ($T_s >$ 600°C, the basic phase is the O phase $Pnma$) transition with $\rho(T) = \rho_{max}$ near the Curie temperature.

(4) The metallic behavior of $\rho(T)$ in the LaSrMnO films studied in this paper ($T_s >$ 600°C) is unrelated to the transition to the metallic state. It is observed when fluctuations are suppressed and the material becomes structurally and magnetically ordered. Under these conditions, the percolation threshold changes and the resistivity diminishes with decreasing temperature.

ACKNOWLEDGMENTS

This work was partially supported by the government of Poland (grant no. PBZ-KBN-013/T08/19).

REFERENCES

1. É. L. Nagaev, Usp. Fiz. Nauk **166**, 833 (1996) [Phys. Usp. **39**, 781 (1996)].
2. O. Yu. Gorbenko, R. V. Demin, A. R. Kaul', *et al.*, Fiz. Tverd. Tela (St. Petersburg) **40**, 290 (1998) [Phys. Solid State **40**, 263 (1998)].
3. Z. L. Wang, J. S. Yin, Y. D. Jang, and Jiming Zhang, Appl. Phys. Lett. **70**, 3362 (1997).
4. V. S. Gaviko, V. E. Arkhipov, A. V. Korolev, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **41**, 1064 (1999) [Phys. Solid State **41**, 969 (1999)].
5. V. G. Prokhorov, G. G. Kaminskiĭ, V. S. Flis, and Li Yang Pak, Fiz. Nizk. Temp. **25**, 1060 (1999) [Low Temp. Phys. **25**, 792 (1999)].
6. I. O. Troyanchuk, O. S. Mantytskaya, S. N. Postushonok, *et al.*, Kristallografiya **41**, 839 (1996) [Crystallogr. Rep. **41**, 797 (1996)].
7. V. N. Strekalovskiĭ, Yu. M. Polezhaev, and S. F. Pal'guev, *Oxides with Impurity Disordering* (Nauka, Moscow, 1987).
8. V. D. Okunev, Z. A. Samoilenko, V. M. Svistunov, *et al.*, J. Appl. Phys. **85**, 7282 (1999).
9. M. A. Krivoglaz, in *Electronic Structure and Electronic Properties of Metals and Alloys* (Naukova Dumka, Kiev, 1998), pp. 3–39.
10. V. D. Okunev and Z. A. Samoĭlenko, Pis'ma Zh. Éksp. Teor. Fiz. **53** (1), 42 (1991) [JETP Lett. **53**, 44 (1991)].
11. K. Meyer, *Physikalisch–Chemische Kristallographie* (VEB, Leipzig, 1968; Metallurgiya, Moscow, 1972).

12. B. I. Shklovskiĭ and A. L. Éfros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer, New York, 1984).
13. V. D. Okunev, Z. A. Samoilenko, A. Abal'oshev, *et al.*, Phys. Rev. B **62**, 696 (2000).
14. V. D. Okunev, N. N. Pafomov, V. A. Isaev, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **44**, 150 (2002) [Phys. Solid State **44**, 157 (2002)].
15. S. G. Kaplan, M. Quijada, H. D. Drew, *et al.*, Phys. Rev. Lett. **77**, 2081 (1996).
16. J. M. D. Coey, M. Viret, L. Ranno, and K. Ounagjela, Phys. Rev. Lett. **75**, 3910 (1995).
17. M. Ziese and C. Srinitiwarawong, Phys. Rev. B **58**, 11519 (1998).
18. I. O. Troyanchuk, L. S. Lobanovskiĭ, D. D. Khalyavin, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 604 (1999) [JETP **89**, 321 (1999)].
19. V. D. Okunev, Z. A. Samoĭlenko, V. A. Isaev, *et al.*, Pis'ma Zh. Tekh. Fiz. **28** (2), 12 (2002) [Tech. Phys. Lett. **28**, 44 (2002)].
20. A. Biswas, M. Rajeswari, R. C. Srivastava, *et al.*, Phys. Rev. B **63**, 184424 (2001).
21. M. B. Salamon and M. Jaime, Rev. Mod. Phys. **73**, 583 (2001).
22. N. F. Mott, *Metal–Insulator Transitions* (Taylor and Francis, London, 1974; Nauka, Moscow, 1979).
23. A. E. Kar'kin, D. A. Shulyatev, A. A. Arsenov, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 671 (1999) [JETP **89**, 358 (1999)].
24. N. G. Bebenin, R. I. Zaĭnullina, V. V. Mashkautsan, *et al.*, Zh. Éksp. Teor. Fiz. **117**, 1181 (2000) [JETP **90**, 1027 (2000)].

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 257–260. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 125–128.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Influence of Laser Irradiation with $h\nu$ = 1.96 eV on the Properties of Titanium–Silicon Contact under Thermal Annealing in Nitrogen

I. S. Nuprienok and A. N. Shibko

El'ton Research and Industrial Commercial Enterprise, Minsk, 220012 Belarus

e-mail: shibko@inel.bas-net.by

Received March 15, 2002

Abstract—The effect of laser irradiation with $h\nu$ = 1.96 eV (λ = 0.63 μm) on the properties of titanium–silicon contacts subjected to steady-state thermal annealing in a nitrogen atmosphere is studied. It is found that treatment-induced changes in the phase composition of the contact modify its electrophysical parameters. The applicability of laser irradiation to the formation of rectifying titanium–silicon junctions with desired parameters is demonstrated. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Refractory metals and their silicides are widely used in microelectronics due to their thermodynamic and electrophysical properties, as well as their high-temperature stability. Modern science and technology require new materials with a particular set of physicochemical properties. In this context, nitrides, compounds of metals with nitrogen, hold much promise for microelectronics. Nitrides offer high hardness, high electrical conductivity, and a high melting point. The superconducting-transition temperature in these compounds is also comparatively high. One method for producing nitride films is steady-state thermal annealing in a nitrogen atmosphere, which, however, is usually accompanied by oxidation. To prevent the oxidation, the annealing is combined with the irradiation of films by electrons of given energies [1, 2]. Recently, much attention has been drawn to nonthermal processes that occur in thin metal films that are simultaneously subjected to thermal treatment and photon irradiation. The photon energy governs the structure, phase composition, and, as a consequence, electrophysical properties of the resulting films [3–5].

In this study, we consider phase transformations and changes in the electrophysical parameters of titanium–silicon compositions subjected to thermal annealing in nitrogen in combination with laser irradiation by photons with an energy $h\nu$ = 1.96 eV.

EXPERIMENTAL

Titanium films with a thickness of 100 nm were prepared by electron-beam deposition at a pressure of 3×10^{-5} Pa on Si(111) *n*-type substrates at a temperature of 100°C. Wafers were pretreated according to the method described in [6]. The films deposited were fine-grained polycrystalline with an average grain size ranging from 15 to 20 nm (Fig. 1a).

The structures thus obtained were scribed into 5 × 5-mm chips, which were placed into a VUP-4 vacuum chamber, where they were simultaneously heat-treated in the nitrogen atmosphere and irradiated by an LGN-215 laser with λ = 0.63 μm. The annealing was carried out at a temperature of 500°C for 1, 5, 15, and 30 min. The laser power was continuously monitored by an IMO-2 power meter and was kept at 55 mW throughout the experiment. Extra-pure-grade nitrogen was frozen out before entering the working chamber. According to our preliminary studies, the annealing at T = 500°C is of most interest. At T < 500°C, phase transformations in the Ti–Si structure are insignificant; at T > 500°C, the film oxidizes, and the processes of oxidation and nitridation are hard to study.

The electron microscopic and electron diffraction analysis of the samples was carried out with a JEM-200CX electron microscope; they were also examined by ESCA. Electron diffraction patterns were interpreted using relevant original studies and ASTM tables. The Schottky barrier height was determined from the *I–V* characteristics as described in [7]. The area of the Ti–Si contact was 0.7 mm^2.

In the absence of laser irradiation, the thermal annealing of the titanium–silicon contacts in a vacuum results in the formation of titanium oxide Ti_3O_5 or titanium silicide TiSi depending on the time of treatment (Table 1, Fig. 1b). Lower titanium oxide forms when titanium combines with oxygen adsorbed on the film during the deposition. According to the data listed in Table 1, titanium reflections that are present on diffraction patterns taken of the samples annealed for 1 min disappear with an increase in the annealing time. Such a behavior is associated with the relationship between

1063-7842/03/4802-0257$24.00

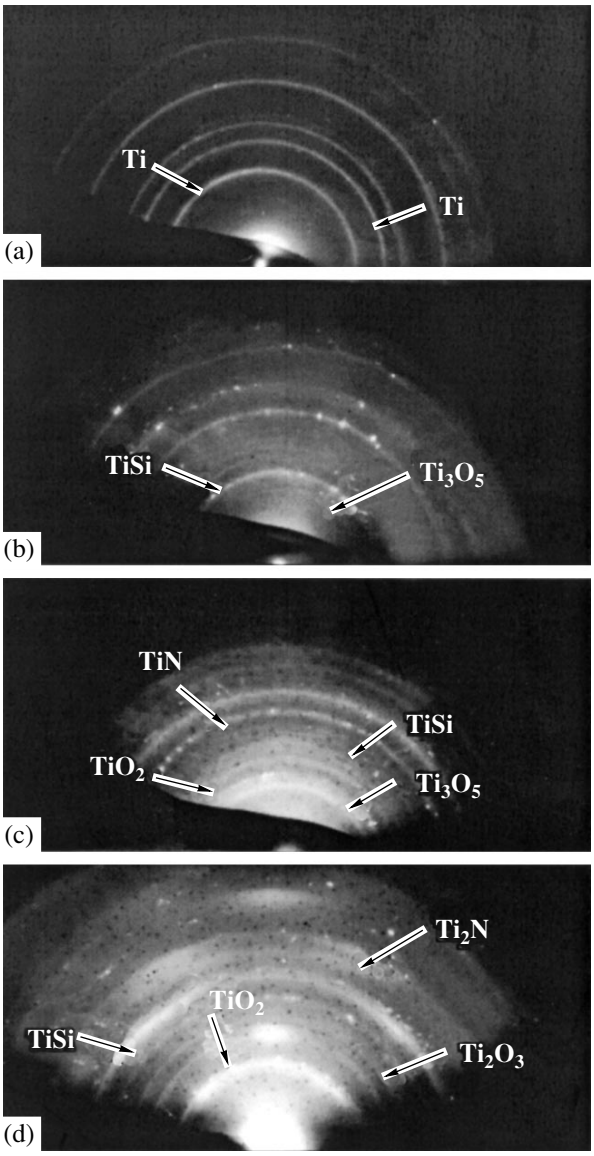


Fig. 1. Electron diffraction patterns from the titanium–silicon compound. (a) Thermal annealing at T = 500°C for 1 min, (b) thermal annealing for 15 min, (c) thermal annealing in extra-pure-grade nitrogen, and (d) annealing combined with laser irradiation (λ = 0.63 μm, E = 55 mW) in the nitrogen.

the rates of Ti–Si interaction and Si oxidation. Reflections from titanium silicide are observed on diffraction patterns taken after the annealing of the Ti–Si structure for τ = 5–30 min. The silicide formation is governed by the diffusion of silicon to the surface and its subsequent interaction with the titanium film. It should be noted that reflections from titanium silicide are particularly intense for τ = 15 and 30 min.

The phase composition of the Ti–Si contact changes under the combined action of thermal annealing in a vacuum (5×10^{-5} Pa) and laser irradiation with λ = 0.63 μm and E = 55 mW. After annealing for 1 min, the diffraction patterns show reflections from only the initial phase (titanium). For annealing times of 5–30 min, reflections due to titanium oxide Ti_3O (the ordered solution of oxygen in titanium) and titanium silicide TiSi arise (Table 2).

Along with the reflections from the titanium oxides and silicide, the annealing of the contact in the atmosphere of extra-pure-grade nitrogen gives rise to reflections from titanium nitrides (Table 3, Fig. 1c). In particular, annealing for 1 min causes reflections from lower oxide Ti_3O_5, initial titanium, and metal-enriched nitride Ti_2N, whereas an increase in the annealing time to 5 min gives rise to reflections from titanium silicide and nitride (TiN). When the annealing time rises to 15–30 min, extra reflections from higher oxide TiO_2 (rutile) also appear on the diffraction patterns.

Special features of metal–nitrogen interaction, the kinetics of saturation by nitrogen, as well as the relation between the nitride layer thickness, amount of absorbed nitrogen, and nitridation time are described in [8]. The nitridation mechanism generally involves diffusion mass transfer of nitrogen into a metal and the formation of the solid solution. Conventional heterodiffusion transforms into reaction diffusion with the formation of nitrides. The nitride formation is limited either by the rate of the reaction itself or by the diffusion mass transfer of nitrogen through the already formed nitride layer [9–11]. Analysis shows that the rate of titanium interaction with nitrogen is considerably lower than that of the interaction with oxygen. This is due to the fact that, at a fixed temperature, the rate of oxygen absorption by titanium is seven times that of nitrogen absorption. The key to this effect evidently lies in the atomic configuration. An oxygen atom has the s^2p^4 configuration and tends to complete it to s^2p^6. A nitrogen atom with the s^2p^3 configuration tends either to complete s^2p^6 shells or to lose an electron, taking the sp^3 configuration. This circumstance also slows down the diffusion of nitrogen as compared with oxygen [10]. Therefore, if there is a tendency toward surface oxidation, an increase in the oxygen content makes nitridation difficult. When combined with metals, nitrogen may either accept their electrons or donate its own electron to form the energetically stable s^2p^6 or sp^3 configuration, respectively. In the former case, the resulting compounds feature pronounced ionic bonding; in the latter, bonds are typically metallic with a greater or lesser share of covalent bonds.

It is known that, under certain conditions, *d* transition Group-IV metals (Ti, Zr, Hf) and Group-V metals (V, Nb, Ta) readily interact with molecular nitrogen to produce nitrides. The temperature coefficient of transition metal nitride formation ranges from 25 to 35 kcal/mol [12].

Thermal annealing in nitrogen in combination with 0.63-μm laser irradiation ($h\nu$ = 1.96 eV) with a power of 55 mW modifies significantly the phase composition

Table 1. Phase composition of the Ti–Si contact after thermal annealing at $T = 500°C$

d, Å	τ, min			
	1	5	15	30
3.54	Ti_3O_5	Ti_3O_5	Ti_3O_5	Ti_3O_5
3.14	Ti_3O_5	Ti_3O_5	Ti_3O_5	Ti_3O_5
2.69	–	TiSi	TiSi	TiSi
2.68	Ti_3O_5	Ti_3O_5	Ti_3O_5	Ti_3O_5
2.44	–	TiSi	TiSi	TiSi
2.35	Ti	–	–	–
2.34	–	TiSi	TiSi	TiSi
2.24	Ti	–	–	–
2.19	–	TiSi	TiSi	TiSi
1.73	Ti	–	–	–
1.48	Ti	–	–	–

Note: *d*, interplanar spacing; τ, annealing time.

Table 2. Phase composition of the Ti–Si contact after thermal annealing at $T = 500°C$ combined with laser irradiation with $\lambda = 0.63$ μm and $E = 55$ mW

d, Å	τ, min			
	1	5	15	30
2.69	–	TiSi	TiSi	TiSi
2.44	–	TiSi	TiSi	TiSi
2.39	–	Ti_3O	Ti_3O	Ti_3O
2.35	Ti	–	–	–
2.34	–	TiSi	TiSi	TiSi
2.24	Ti	–	–	–
2.19	–	TiSi	TiSi	TiSi
2.17	–	Ti_3O	Ti_3O	Ti_3O
1.73	Ti	–	–	–
1.68	–	Ti_3O	Ti_3O	Ti_3O
1.48	Ti	–	–	–
1.42	–	Ti_3O	Ti_3O	Ti_3O

Table 3. Phase composition of the Ti–Si contact after thermal annealing at $T = 500°C$ in extra-pure-grade nitrogen

d, Å	τ, min			
	1	5	15	30
3.54	Ti_3O_5	Ti_3O_5	Ti_3O_5	Ti_3O_5
3.28	–	–	TiO_2	TiO_2
3.24	–	–	TiO_2	TiO_2
3.14	Ti_3O_5	Ti_3O_5	Ti_3O_5	Ti_3O_5
2.69	Ti	–	–	–
2.68	Ti_3O_5	Ti_3O_5	Ti_3O_5	Ti_3O_5
2.56	–	TiSi	TiSi	TiSi
2.48	–	TiN	TiN	TiN
2.44	–	TiSi	TiSi	TiSi
2.34	Ti	–	–	–
2.32	Ti_2N	Ti_2N	Ti_2N	Ti_2N
2.25	Ti_2N	Ti_2N	Ti_2N	Ti_2N
2.24	Ti	–	–	–
2.19	–	TiSi	TiSi	TiSi
2.16	–	–	TiO_2	TiO_2
2.09	–	TiN	TiN	TiN
1.76	Ti_2N	Ti_2N	Ti_2N	Ti_2N
1.47	–	TiN	TiN	TiN

Table 4. Phase composition of the Ti–Si contact after thermal annealing at $T = 500°C$ combined with laser irradiation with $\lambda = 0.63$ μm and $E = 55$ mW in extra-pure-grade nitrogen

d, Å	τ, min			
	1	5	15	30
3.28	–	–	TiO_2	TiO_2
3.24	–	–	TiO_2	TiO_2
2.71	–	Ti_2O_3	Ti_2O_3	–
2.69	–	TiSi	TiSi	TiSi
2.56	Ti	Ti	–	–
2.48	–	–	–	TiN
2.44	–	TiSi	TiSi	TiSi
2.34	Ti	Ti	–	–
2.33	–	Ti_3O	Ti_3O	Ti_3O
2.32	–	–	Ti_2N	Ti_2N
2.25	–	–	Ti_2N	Ti_2N
2.24	Ti	Ti	–	–
2.19	–	TiSi	TiSi	TiSi
2.18	–	Ti_3O	Ti_3O	Ti_3O
2.16	–	–	TiO_2	TiO_2
2.09	–	–	–	TiN
1.76	–	–	Ti_2N	Ti_2N
1.70	–	Ti_2O_3	Ti_2O_3	–
1.47	–	–	–	TiN

of the titanium–silicon contact. It is seen from Table 4 that short-term annealing (1 and 5 min) suppresses oxidation and nitridation. An increase in the annealing time to τ = 15 and 30 min results in the formation of TiN (Fig. 1d) and Ti_2N, respectively. The absence of lower high-oxygen oxide Ti_3O_5 and titanium nitrides after annealing for 1 and 5 min can be attributed to the fact that the irradiation of the composition being

Table 5. Electrophysical parameters of the titanium–silicon contact

T, °C	τ, min	φ, eV	*U*, V	*n*
As-prepared sample		0.55	6	1.17
	After thermal annealing			
500	1	0.56	7	1.16
	5	0.56	8	1.16
	15	0.56	9	1.16
	30	0.57	15	1.12
	After thermal annealing combined with laser irradiation			
500	1	0.56	8	1.16
	5	0.56	8	1.16
	15	0.56	9	1.09
	30	0.57	12	1.07
	After thermal annealing in extra-pure-grade nitrogen			
500	1	0.54	8	1.17
	5	0.54	8	1.17
	15	0.53	10	1.18
	30	0.52	12	1.19
	After thermal annealing combined with laser irradiation in extra-pure-grade nitrogen			
500	1	0.56	8	1.16
	5	0.56	8	1.16
	15	0.56	10	1.17
	30	0.54	12	1.18

Note: φ is the Schottky barrier height; *U*, breakdown voltage; and *n*, ideality coefficient.

annealed by 1.96-eV photons suppresses (at a certain photon flux density) oxidation and nitridation and the dissolved oxygen and nitrogen become chemically inactive [13]. For the complete suppression of oxidation and nitridation in the titanium–silicon contact during heat treatment in a nitrogen atmosphere, the output of 0.63-μm laser radiation (photons with $h\nu = 1.96$ eV) should be increased in order to passivate M–O and M–N bonds and thereby prevent the interaction of oxygen and nitrogen with the contact surface (titanium film) during the annealing. The electron diffraction and ESCA data correlate.

The change in the phase composition of Ti–Si entails a change in the electrophysical parameters of the contact. The *I–V* characteristic of the as-prepared contact was asymmetric with a breakdown voltage of 6 V, which indicates the presence of silicon dioxide SiO_2 between the silicon substrate and titanium film. The electrophysical parameters of the contact after various treatments are summarized in Table 5. The annealing decreases the concentration of defects in the insulating SiO_2 film at the metal–silicon interface. The diffusion of oxygen thickens the insulating layer, as evidenced by the increase in the breakdown voltage. The Schottky barrier grows after thermal annealing in a vacuum and diminishes after annealing in a nitrogen atmosphere. The treatment alters the density of surface states in the silicon at the metal–silicon interface, thereby changing the barrier height. The *I–V* characteristics of the contact were characterized by the ideality coefficient *n*, which was estimated from the experimental dependences [7] (Table 5). It is demonstrated that the ideality coefficient decreases with increasing vacuum annealing time and grows after annealing in nitrogen. This is associated with the modification of the electrophysical parameters of the oxide layer at the interface and with a treatment-induced change in the concentration of electron traps. Therefore, the electrophysical parameters of the titanium–silicon contact are sensitive to its phase composition, the formation of titanium nitrides and oxides, and the variation of the interface constituents. The initial $Ti/SiO_2/Si$ interface transforms into $Ti_xO_y/TiSi$, Ti_xO_y, Ti_mN, or $TiSi/SiO_2/Si$ depending on the type of treatment.

Thus, it is demonstrated that the combined processing of the titanium–silicon composition in nitrogen makes it possible to form a rectifying contact with a given phase composition and, as a consequence, with a given set of electrophysical properties.

REFERENCES

1. D. T. Alimov, V. K. Tyugaĭ, P. K. Khabibulaev, *et al.*, Zh. Fiz. Khim. **61**, 3065 (1987).
2. A. M. Chaplanov and A. N. Shibko, Izv. Akad. Nauk SSSR, Ser. Fiz. **53**, 1111 (1989).
3. W. Ian and I. W. Boyd, Appl. Surf. Sci. **109–110**, 538 (1997).
4. I. Y. Zhand and I. W. Boyd, Appl. Phys. Lett. **71**, 2964 (1997).
5. I. S. Nuprienok and A. N. Shibko, Zh. Tekh. Fiz. **71** (9), 45 (2001) [Tech. Phys. **46**, 1112 (2001)].
6. *Handbook of Thin Technology*, Ed. by L. I. Maissel and R. Glang (McGraw-Hill, New York, 1970; Sovetskoe Radio, Moscow, 1977), Vol. 1.
7. E. H. Rhoderick, *Metal–Semiconductor Contacts* (Clarendon, Oxford, 1978; Radio i Svyaz', Moscow, 1982).
8. S. S. Kiparisov and Yu. V. Levitskiĭ, *Nitridation of Refractory Metals* (Metallurgiya, Moscow, 1972).
9. H. Sumi, H. Inoue, and Y. Sugano, Jpn. J. Appl. Phys., Part 1 **36**, 595 (1997).
10. H. Kheyrandish, J. S. Colligan, and J. K. Kim, J. Vac. Sci. Technol. A **12**, 2723 (1994).
11. R. A. Andrievskiĭ, G. M. Vol'dman, and M. A. Leont'ev, Izv. Akad. Nauk SSSR, Neorg. Mater. **27**, 729 (1991).
12. M. D. Lyutaya, *Production, Properties, and Application of Nitrides* (Kiev, 1972), pp. 6–13.
13. A. M. Chaplanov and A. N. Shibko, Neorg. Mater. **29**, 1477 (1993).

Translated by A. Sidorova

Technical Physics, Vol. 48, No. 2, 2003, pp. 261–264. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 129–132.

BRIEF COMMUNICATIONS

Inhomogeneous $SiO_x\langle Fe\rangle$ Metal-Dielectric Films as a Material for Infrared Thermal Radiation Detectors

E. V. Michailovskaya, I. Z. Indutnyy, and P. E. Shepeliavyi

Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, pr. Nauki 45, Kiev, 03028 Ukraine

e-mail: indutnyy@isp.kiev.ua

Received March 19, 2002

Abstract—A vacuum deposition method for producing $SiO_x\langle Fe\rangle$ composite metal-dielectric films in which the metal concentration in the SiO_x dielectric matrix varies across thickness is described. The reflection and transmission of the films in the 2–12 μm IR spectral range are studied and their optical properties are simulated. In the temperature range 283–390 K, the temperature-sensitive properties of the $SiO_x\langle Fe\rangle$ films with a phase volume ratio of 23% (Fe) : 77% (SiO) is investigated. For these films, the temperature coefficient of resistance is found. The feasibility of these films as a sensitive layer in microbolometers is demonstrated. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In recent years, much attention has been concentrated on the design of small thermal imagers and IR displays where the temperature-sensitive element is made in the form of a chain or array consisting of identical uncooled thermal radiation detectors (TRDs) [1]. Their basic advantage is the absence of a system for cryogenic cooling, which significantly facilitates the design of IR displays, reduces their weight and cost, etc. In contrast to photon detectors, TRDs can operate in a wide IR spectral range, since their long-wave sensitivity is unlimited. The family of TRDs includes thermocouples, bolometers, and acoustooptic and pyroelectric detectors. Array bolometers are the easiest to fabricate; therefore, they are fairly popular. Normally, the sensitive element in a microbolometer consists of a temperature-sensitive layer with a high value of the temperature coefficient of resistance (TCR) and an absorbing coating. As the temperature-sensitive layer, VO_2 films [2, 3] or amorphous semiconductor films [4, 5] are most often used. As the absorbing coating, one can apply thin metal films (an absorption of ≈50% in a wide spectral region) or multilayer structures for which the resonant absorption may reach 90% in a certain spectral range.

However, these additional layers raise the heat capacity of a microbolometer substantially and adversely affect its performance. Therefore, searching for materials that would combine the functions of temperature-sensitive and absorbing layers seems to be topical. It was shown earlier [6] that inhomogeneous metal-dielectric films can offer a high absorption in the IR range combined with elevated (in comparison with metals) TCR values. This makes it possible to accomplish the absorption of IR radiation and the formation of an electrical signal in a single layer.

In this study, we investigate the optical and thermoresistive properties of inhomogeneous $SiO\langle Fe\rangle$ metal-dielectric films. Experimental data for reflection (R) and transmission (T) spectra for these films at wavelengths between 2 and 12 μm are reported. These spectra are compared with those calculated in the effective-medium approximation. In addition, the temperature dependences of the film resistance (r) and TCR in the interval from 283 to 390 K are presented.

EXPERIMENTAL

Samples were prepared by the method proposed earlier to produce light-absorbing inhomogeneous SiO- and Cr-based coatings [7]. Its basic idea is the thermal coevaporation of a mixture of these materials in a vacuum from a single evaporator. In this work, a mixture of SiO and Fe powders was deposited on glass, pyroceramic, and silicon substrates at a pressure of 2×10^{-3} Pa. The average metal-to-oxide mass ratio in a layer thus obtained was roughly determined from the ratio of the powder masses in the evaporator, and the distribution of the components was specified by the evaporation regime. Since SiO and Fe evaporate at different temperatures, it is possible to select such conditions for mixture heating that would provide a desirable distribution of the component concentrations in the layer deposited. For example, a gradual decrease in the temperature of the evaporator leads to a continuous decrease in the metal concentration in the vapor flux. Layers thus obtained are composite $SiO_x\langle Fe\rangle$ films in which the metal concentration in the dielectric matrix varies with thickness. Note that the concentration of the metal is maximal near the interface and minimum at the film

1063-7842/03/4802-0261$24.00

surface. Such a distribution specifies the optical and electrical properties of these inhomogeneous metal-dielectric layers. The deposition process was controlled by a calibrated quartz resonator. The thickness of the films measured by a microinterferometer was found to range between 2 and 3 μm.

Layers obtained by the thermal evaporation of silicon oxide under similar conditions had the composition SiO_x, where $x = 1.2–1.3$ [8]. Therefore, hereafter we designate the films under consideration as $SiO_x\langle Fe\rangle$. The reflection and transmission spectra in the IR spectral range were taken from the films deposited on Si(100) substrates using an IKS-31 spectrometer. To examine the temperature-sensitive properties (TCR and resistance), we used the films deposited on the pyroceramic substrates in a vacuum. Samples had the form of a planar chip of several square millimeters in area. On the top of the chip, Mn–Ag contacts were deposited to which thin wires were soldered. During the measurements, the samples were placed in a thermostat and the bias current was chosen in such a way as to prevent Joule heating.

EXPERIMENTAL RESULTS AND DISCUSSION

Figures 1 and 2 show the measured spectral dependences of R and T for 2.73-μm-thick $SiO_x\langle Fe\rangle$ films. The powder mass ratio in the evaporator was 50 : 50, which corresponded to a Fe-to-SiO volume ratio of 23 : 77. The deposition conditions provided a nonuniform distribution of the phases in the film with the minimum concentration of the metal on its surface. As was shown earlier [9], such a distribution ensures a low reflectivity in a wide spectral region. As follows from Fig. 2, in the entire spectral region considered, the film is characterized by a low transmission (less than 1%). Therefore, its absorptivity equals $1 - R$. In the range of 2–8.5 μm, the absorptivity is about 90%. In the longer wave range (8.5–12 μm), the reflectivity rises and peaks (32.5%) near 9.5 μm. Such behavior of the reflection spectrum is associated with the presence of the absorption band (typical of SiO_x) due to the Si–O bond vibration mode [10]. This band is also observed in the transmission spectrum (Fig. 2). However, in this range, too, the film has a high absorptivity, which ranges from 85 to 67.5%.

The simulation of the optical characteristics for these layers in the IR region was performed using the method [9] applied to the visible range. It was assumed that a film deposited can be represented as a metal-dielectric layer consisting of an oxide matrix with metal inclusions whose concentration varies linearly across the layer; the average volume concentration of the metal is 23% (i.e., the same as in the evaporator). To describe the optical constants (refractive index n and absorption coefficient k) of such a composite, we used the Bruggeman–Landauer theory of effective dielectric function of statistical heterostructure [11, 12]. Since there are no detailed experimental data for the optical constants of SiO_x (for $x = 1.2–1.3$) in the IR region, we applied the data for n and k of SiO [13] to the matrix. For iron, n and k were taken from [14]. The spectral dependences of n and k obtained for composites with different Fe concentrations were used to simulate the reflection and transmission of the film deposited. The parameters R and T of an inhomogeneous layer were calculated by applying the Abeles matrix theory, which is stated for dielectric and inhomogeneous nonabsorbing structures in [15] and [16], respectively (for details of the calculation technique, see [9]). Since the Fe distribution in the layer was not studied experimentally and was specified by the deposition conditions, we calculated R and T of the layer for different gradients of the metal concentration, keeping the average concentration constant. The discrepancy between the experi-

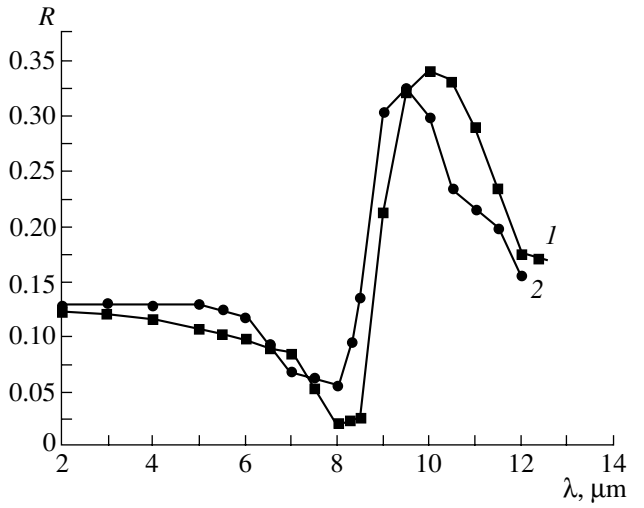


Fig. 1. Reflection spectrum of the $SiO_x\langle Fe\rangle$ gradient layer. (■) Calculation and (●) experimental data.

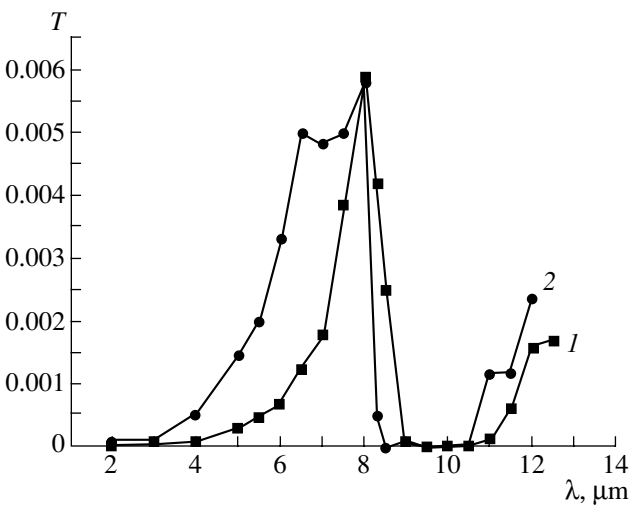


Fig. 2. Transmission spectrum of the $SiO_x\langle Fe\rangle$ gradient layer (designations are the same as in Fig. 1).

mental and simulated spectra was minimized by selecting the appropriate gradient.

Figures 1 and 2 show the optimum R and T spectra calculated for the 2.73-μm-thick gradient $SiO\langle Fe\rangle$ layer. These curves correspond to a linear increase in the iron volume fraction in the layer from 42% (near the interface) to 4% (on the layer surface). Agreement between the calculated and experimental curves is satisfactory, indicating that the model of an inhomogeneous metal-dielectric layer is applicable in our case. However, there are certain differences associated with the approximations and simplifications used in the calculations. First, the structure related to the vibrational band in the simulated reflection spectrum shifts toward the long-wave range as compared with the experiment. This is because we used the optical data for SiO rather than for SiO_x. As is known [10], the position of the vibrational band maximum depends on the stoichiometric ratio and shifts toward the short-wave region with increasing x, which is the case. In addition, the bands are slightly distorted, especially in the transmission spectrum. It seems likely that SiO and Fe react in the evaporator during the evaporation to form a certain amount of iron silicides, which may alter the matrix composition and cause extra bands in the IR region to appear.

In general, when choosing a material for the temperature-sensitive layer of a microbolometer, we should be guided by its resistance, TCR, and specific heat [1]. The investigation of the electrophysical properties of the metal-dielectric films showed that their resistivity depends essentially on the composition. By varying the metal content in the film, one can vary its resistivity in a very wide range from values typical of a dielectric oxide to those of a metal. The resistance of the inhomogeneous layers is specified mainly by the region where the metal concentration is maximal. However, it also depends on the concentration gradient. A change in the volume ratio between the conducting (Fe) and nonconducting (SiO_x) phases, as well as the particle size distribution, significantly influences the TCR value [16].

To determine the TCR in the inhomogeneous $SiO_x\langle Fe\rangle$ films, we recorded the temperature dependences of their resistance r. Test samples were 0.8 × 2-mm chips on pyroceramic substrates. By varying the metal content (the mean concentration was varied from 13 to 23 vol%) and setting the appropriate evaporation conditions for the SiO–Fe mixture, we succeeded in growing films with a TCR varying within (0.3–1.2)% K^{-1} (at 293 K). The resistance of the chips was from several hundreds of ohms to several hundreds of kiloohms, which falls into the interval of optimum resistances for microbolometers [1]. The maximum TCR values were observed for the highest resistivity samples.

For such samples, a typical temperature dependence of the resistance r constructed in the $\ln r - T^{-1}$ coordinates is shown in Fig. 3. The results presented were obtained for an inhomogeneous $SiO_x\langle Fe\rangle$ film with the same parameters as in Figs. 1 and 2 (a thickness of 2.73 μm and a mean Fe-to-SiO volume ratio of 23:77). It is seen that the resistance drops with temperature (i.e., the films have a negative TCR) and its room-temperature value is equal to –0.47% K^{-1}, decreasing slightly with temperature (Fig. 3). The negative TCR value is likely to be associated with the conduction mechanism in these films (tunnel conduction through percolation channels as, for example, in Au–SiO cermet films of similar composition, for which this mechanism has been studied rather extensively [16, 17]).

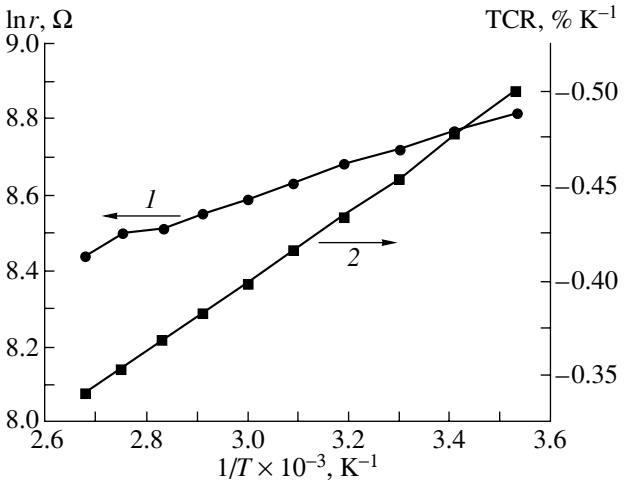


Fig. 3. Temperature dependences of the (●) resistance and (■) TCR of the $SiO_x\langle Fe\rangle$ film measuring 0.8 mm × 2 mm × 2.73 μm.

CONCLUSION

Thus, using thermal evaporation in a vacuum, we obtained inhomogeneous $SiO_x\langle Fe\rangle$ metal-dielectric films, which are characterized by a low reflection coefficient and high absorption coefficient in the near- and mid-IR spectral ranges. The optical parameters of these films are described by the model of the effective dielectric function of a statistical oxide matrix–metal inclusion heterosystem where the concentration of metal inclusions across the film obeys a linear law. The electrical resistance and TCR of the films, as well as their optical properties, depend on the concentration of metal inclusions and concentration gradient. In particular, for films with the average component volume ratio Fe : SiO = 23:77, the TCR equals –0.47% K^{-1} at room temperature. Such inhomogeneous metal-dielectric layers can be used in microbolometers as sensitive layers that combine IR radiation absorption and electrical signal formation.

REFERENCES

1. I. A. Khrebtov and V. G. Malyarov, Opt. Zh. **64** (6), 3 (1997) [J. Opt. Technol. **64** (6), 511 (1997)].

2. V. Yu. Zerov, Yu. V. Kulikov, V. N. Leonov, *et al.*, Opt. Zh. **66** (5), 8 (1999) [J. Opt. Technol. **66** (5), 387 (1999)].
3. Changhong Chen, Xinjian Yi, Jing Zhang, and Xingrong Zhao, Infrared Phys. Technol. **42**, 87 (2001).
4. K. C. Liddiard, Infrared Phys. **26** (1), 3 (1986).
5. K. C. Liddiard, M. H. Unewisse, and O. Reinhold, Proc. SPIE **2225**, 62 (1994).
6. P. E. Shepeliavyi, K. V. Michailovskaya, and I. Z. Indutnyi, in *Proceedings of International Conference on Advanced Materials* (Kiev, 1999), p. 52.
7. P. E. Shepeliavyi, E. V. Michailovskaya, I. Z. Indutnyy, *et al.*, Opt. Tekh., No. 2 (6), 16 (1995).
8. I. Ya. Bratus', V. A. Yukhimchuk, L. I. Berezhinskiĭ, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **35**, 854 (2001) [Semiconductors **35**, 821 (2001)].
9. I. Z. Indutnyy, P. E. Shepeliavyi, E. V. Michailovskaya, *et al.*, Zh. Tekh. Fiz. **72** (6), 67 (2002) [Tech. Phys. **47**, 720 (2002)].
10. M. Nakamura, Y. Mochizuki, K. Usami, *et al.*, Solid State Commun. **50**, 1079 (1984).
11. D. A. G. Bruggeman, Ann. Phys. (Leipzig) **24**, 636 (1935).
12. R. Landauer, J. Appl. Phys. **23**, 779 (1952).
13. *Handbook of Optical Constants of Solids*, Ed. by. E. D. Palik (Academic, London, 1985), p. 805.
14. P. B. Johnson and R. W. Christy, Phys. Rev. B **9**, 5056 (1974).
15. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon, Oxford, 1969; Nauka, Moscow, 1970).
16. R. Jacobson, in *Physics of Thin Films: Advances in Research and Development* (Academic, New York, 1975; Mir, Moscow, 1978), Vol. 8, pp. 61–105.
17. J. E. Morris, Thin Solid Films **11**, 299 (1972).

Translated by Yu. Vishnyakov

Technical Physics, Vol. 48, No. 2, 2003, pp. 265–266. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 133–135.

BRIEF COMMUNICATIONS

Electroluminescent Sources of Light Based on ZnSe : Mn Single Crystals with Optimum Brightness

M. F. Bulanyĭ, A. V. Kovalenko, and B. A. Polezhaev

Dnepropetrovsk National University, pr. Gagarina 72, Dnepropetrovsk, 49050 Ukraine

e-mail: bulanui@mail.dsu.dp.ua

Received April 19, 2001

Abstract—Electroluminescent sources of light based on plastically deformed ZnSe : Mn single crystals are studied. It is shown that the emissivity of the electroluminescent sources depends on the manganese concentration and the part of a ZnSe : Mn boule from which a plane-parallel wafer is cut. *© 2003 MAIK "Nauka/Interperiodica".*

It is known that zinc sulfide and zinc selenide are widely used in fabricating electroluminescent sources [1–3]. Among zinc sulfide and zinc selenide phosphors, ZnS : Mn and ZnSe : Mn single crystals occupy an important place. Waters for electroluminescent sources are cut from single-crystal ZnSe : Mn boules [4]. To increase the brightness of the source, $2 \times 2 \times 4$-mm wafers cut from a ZnSe : Mn (111) single-crystal boule at an angle of $45° \pm 2°$ to the (111) plane are subjected to uniaxial compression to produce plastic strains of 2–5% [5].

Our studies have shown that the emissivity of a light source thus prepared depends on the part of a single-crystal ZnSe : Mn boule from which the wafer is cut. The scatter in the brightness of the wafers cut from the same single crystal with a fixed plastic strain ($\varepsilon \approx 5\%$) and manganese concentration in the starting mixture ($\approx 5 \times 10^{-2}$ wt %) reaches 60% (when the brightness of the wafers cut from the bottom and center parts are compared) and 40% (for the wafers cut from the top and center parts). For a manganese concentration of $\approx 5 \times 10^{-2}$ wt % in the wafer, the scatter in the electroluminescence intensity between the wafers cut from the same portions of a single-crystal boule was ≈20 and 8%, respectively. As will be shown below, the reason for such a spread is the nonuniform distribution of the manganese dopant along a single-crystal zinc selenide boule grown from the melt under argon pressure.

The purpose of this work is to study conditions for fabricating electroluminescent sources with optimum brightness. To find reasons for the brightness scatter in wafers cut from the same single-crystal boule, EPR analysis of the manganese distribution along a single-crystal ZnSe : Mn boule was carried out. The analysis showed that, as a whole, the boule was doped nonuniformly. In its central part (about 70–75% of the length), the manganese distribution was found to be uniform. If the concentration in this part is conventionally taken to be 100%, then in the bottom part (up to 15% of the length), which crystallizes first, the manganese concentration decreases to about 50–70%. In the top part of the boule (15% of the length), which is the last to grow, the manganese concentration is, on the contrary, higher, rising to 130–150%. Such a distribution of the dopant along a boule grown from the melt under argon pressure could be expected, because single crystal growth in this case involves the same processes as in zone refining of materials. As the solubility of an impurity in the melt is higher than in the solid, at the initial stage of growth, impurities present in the charge pass from the solid phase to the melt zone. Consequently, the bottom portion of the crystal (which crystallizes first) will contain a lesser amount of impurity than the central portion, which, in turn, will be less doped than the top portion of the boule (which is the last to grow).

It can be assumed that the effect of plastic deformation on the brightness depends on the manganese concentration in the samples. To check this assumption, the brightness of equally strained samples was studied as a function of the manganese concentration. Samples from the same lot and from different lots were studied. The manganese concentration was determined by two independent methods: EPR and chemical analysis. The electroluminescence brightness and resistivity of the equally strained ZnSe : Mn single crystals vs. the manganese concentration in the samples are listed in the table.

The results indicate that the emissivity of the working media prepared by the procedure described in [5] varies with the manganese concentration. Since the initial brightness depends on the manganese concentration, the table lists the relative changes in the brightness of the strained samples at a fixed manganese concentration; i.e., the effect of the plastic strain on the brightness is isolated. We plotted the electroluminescence brightness for a fixed plastic strain versus manganese concentration in the samples studied (see figure). The plot has two regions: a linear one, where the brightness varies in

1063-7842/03/4802-0265$24.00

Electroluminescence brightness and resistivity of plastically strained ZnSe vs. the dopant (manganese) concentration

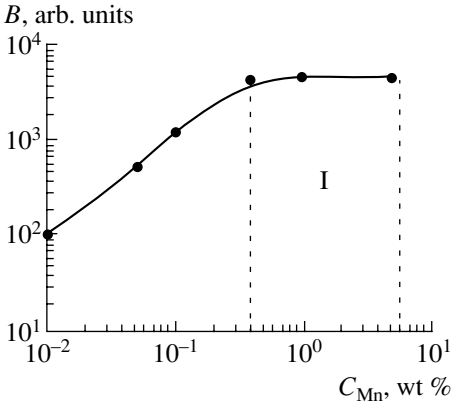
Plastic strain ε = 5%						
Mn concentration, wt %	10^{-2}	5×10^{-2}	10^{-1}	5×10^{-1}	1	5
B	1.1×10^{2}	0.9×10^{3}	1.5×10^{3}	2.0×10^{3}	3.5×10^{3}	4.1×10^{3}
ρ	2.8×10^{-8}	4.2×10^{-7}	7.5×10^{-7}	1.3×10^{-7}	8.0×10^{-8}	4.5×10^{-8}

Note: *B* is the relative increase in brightness in the strained sample in comparison with the as-grown one and ρ is the relative decrease in the resistivity of the strained sample in comparison with the as-grown one.

proportion to the concentration of manganese ions in the working medium, and a region of saturation, where the brightness depends on the concentration of manganese ions only slightly. The latter is observed at high dopant concentrations.

In order to reduce the scatter in the brightness of the plastically strained working media and also to pick out the most efficient emitters, we set the lower limit of the Mn ion concentration in the working medium at a level of 5×10^{-1} wt %. Starting from this concentration, the linear portion of the electroluminescence–Mn the concentration curve changes to the saturation region.

At Mn concentrations $>5 \times 10^{-1}$ wt %, the brightness curve of the plastically strained working media runs nearly parallel to the concentration axis. This ensures a high brightness with a minimum scatter from sample to sample.



Brightness of electroluminescent sources fabricated from ZnS : Mn single-crystal wafers plastically strained to ε = 5% vs. Mn concentration *C*. I is the region of the highest brightness.

The upper boundary of the Mn ion concentration equals 5 wt %. Above this value of the activator concentration, Mn ions in a single-crystal ZnSe boule do not all represent substitutional or interstitial point centers, because Mn precipitates either as an additional solid phase (metallic manganese) or as the MnSe compound, producing the ternary compound $Zn_xMn_{1-x}Se$. The formation of the additional solid phase depends strongly on the single crystal growth conditions, chemical composition of the activator, and presence of coactivators. Wafers containing solid precipitates of manganese or its compounds are harder to be plastically strained. Because of their inhomogeneity, they tend to crack during the treatment; therefore, the sequence of operations used in [5] may be inappropriate in this case.

Thus, the results obtained will help to produce working media for electroluminescent sources with maximum brightness, reproducible characteristics, and a lesser scatter from sample to sample.

REFERENCES

1. A. N. Gorgobiani and P. A. Pipinis, *Tunnel Phenomena in Fluorescence of Semiconductors* (Mir, Moscow, 1994).
2. N. T. Gurin, A. V. Shlyapin, and O. Yu. Sabitov, Zh. Tekh. Fiz. **72** (2), 74 (2002) [Tech. Phys. **47**, 215 (2002)].
3. M. F. Bulanyĭ, B. A. Polezhaev, and T. A. Prokof'ev, Zh. Tekh. Fiz. **67** (10), 132 (1997) [Tech. Phys. **42**, 1229 (1997)].
4. M. F. Bulanyĭ, V. I. Klimenko, and B. A. Polezhaev, Neorg. Mater. **32** (1), 26 (1996).
5. S. A. Omel'chenko, P. A. Berlov, S. I. Bredikhin, *et al.*, Fiz. Tverd. Tela (Leningrad) **24**, 2803 (1982) [Sov. Phys. Solid State **24**, 1586 (1982)].

Translated by B. Kalinin

Technical Physics, Vol. 48, No. 2, 2003, pp. 267–269. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 136–138.

BRIEF COMMUNICATIONS

Effective Nonchain HF Laser Excited by a Self-Sustained Discharge

A. N. Panchenko, V. M. Orlovskiĭ, and V. F. Tarasenko

Institute of High-Current Electronics, Siberian Division, Russian Academy of Sciences, Akademicheskiĭ pr. 4, Tomsk, 634055 Russia

e-mail: VFT@loi.hcei.tsc.ru

Received May 6, 2002

Abstract—Lasing and discharge in mixtures of SF_6 with hydrogen and hydrocarbons are investigated. If the chemical reaction is initiated by a self-sustained discharge (a generator with both an inductive and capacitive energy storage can be used), the maximum lasing efficiency is attained at high values of the ratio E/p (where E is the electric field intensity and p is the pressure in the mixture) if shaped electrodes and preionization of the discharge gap are employed. For the first time, the lasing efficiency obtained in a nonchain HF laser is as high as 10% of the energy deposit and 4.5% of the stored energy for a specific radiation energy of ≈140 J/(l atm). At high efficiencies, the emission spectrum of the nonchain HF laser significantly broadens and cascade lasing at $P(3–2) \longrightarrow P(2–1) \longrightarrow P(1–0)$ vibrational transitions for several rotational lines is realized. © *2003 MAIK "Nauka/Interperiodica".*

(1) Interest in HF and DF lasers pumped by nonchain chemical reactions that are initiated by single-pulse and pulse–periodic self-sustained discharges is related to their high output in the IR spectral range [1–9]. However, the efficiency of an electric discharge nonchain HF laser with respect to the stored energy is usually no greater than 3–4%, and the specific radiation energy is about 70 J/(l atm). If a nonchain HF laser is excited by an inductive-storage generator, the lasing efficiency amounts to 5.5%, including at relatively high specific energies (≈1.7 kJ/(l atm)) [3, 7]. When a pumping discharge is initiated by X rays at E/p exceeding the static breakdown value, the lasing efficiency in mixtures with H_2 relative to the energy deposit is about 8% at relatively low specific energies (about 0.2 kJ/(l atm) without considering the pressure of buffer neon) [4]. Recently [7], we reported a high output energy of an HF laser when using inductive energy storage and intense preionization in mixtures with hydrocarbons and hydrogen. We also proposed the use of shaped electrodes to maximize the output of such lasers.

In this work, we gained effective pumping of a nonchain HF laser by using inductive and capacitive energy storage. For the first time, the efficiency of an electric discharge HF is as high as ≈10% of the energy deposit. The specific radiation energy is about 140 J/(l atm), and the efficiency with respect to the stored energy is 4.5%.

(2) In experiments, we used an inductive-storage laser and a semiconductor current interrupter described in detail elsewhere [9]. The gap between 70-cm-long shaped electrodes was 2 or 3.8 cm. In both cases, the lasing volume was 200 cm^3. Preionization by 72 spark gaps evenly distributed on both sides of the anode ensures the uniform irradiation of the cathode. The value of peaking capacitors was 2.2 nF, and the storage capacitance was decreased to 70 nF. The duration of a discharge current pulse was about 100 ns. The inductive storage can easily be switched off. In this case, the pumping pulse duration increased to 200 ns. Lasing was studied in SF_6–pentane and SF_6–hydrogen mixtures at pressures ranging from 0.02 to 0.2 atm.

(3) The output of the laser with a 2-cm gap was no higher than 0.4 J both with and without the inductive storage. The discharge was heavily inhomogeneous. If the inductive storage is switched off, discharge contraction may take place in 100 ns after gap breakdown.

The laser parameters were improved significantly by extending the gap to 4 cm. With the capacitive storage, the output exceeded 1.0 and 1.4 J in the mixtures with hydrogen and pentane, respectively, when the efficiency with respect to the stored energy was 4.5% and the specific energy approached 140 J/(l atm). As in [7], a residual voltage was detected across the gap after the termination of the current pulse. However, the use of the shaped electrodes and uniform preionization of the cathode prevented the secondary breakdown of the discharge gap. The residual energy in the capacitor evolved through the charging resistance for a long time (several hundreds of microseconds).

With the inductive storage, the output was no greater than 0.9 J because of the nonoptimal operation of the current interrupter. We observed a totally homogeneous discharge and many small bright spots on the cathode (the overall pattern of the discharge glow was recorded by a digital camera).

1063-7842/03/4802-0267$24.00 © 2003 MAIK "Nauka/Interperiodica"

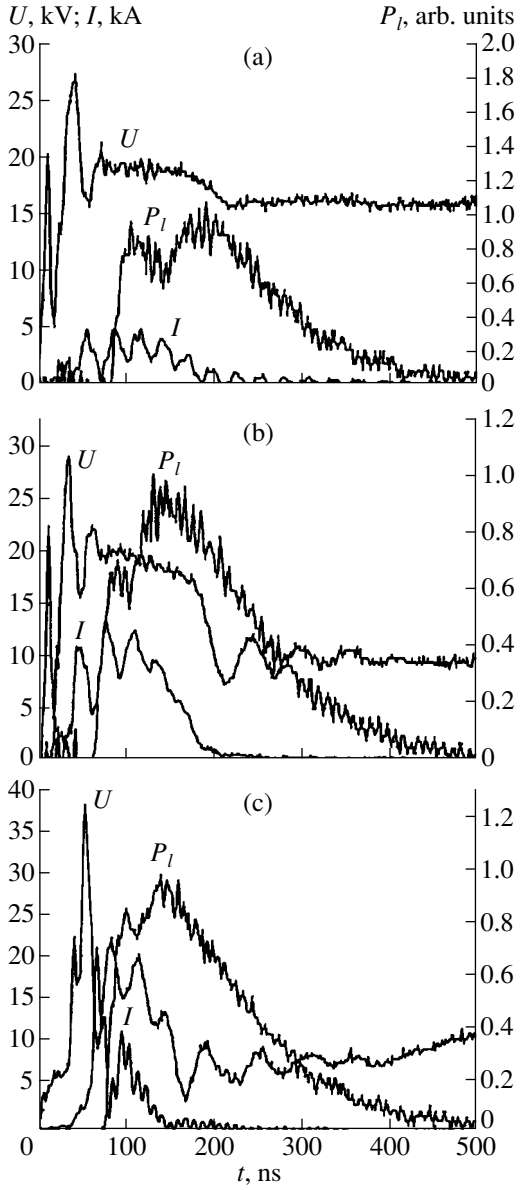


Fig. 1. Voltage, current, and laser output waveforms at U_0 = (a) 22 and (b, c) 30 kV for (a, b) capacitive and (c) inductive pumping. $SF_6 : C_5H_{12} = 18 : 1$ mixture, the pressure p = 0.05 atm.

Figure 1 shows the waveforms of voltage pulses, discharge current, and HF laser output for various working conditions. In the case of capacitive pumping, the residual voltage increased with decreasing charge voltage U_0 and the discharge was not initiated starting from a certain voltage depending on the mixture composition and pressure (e.g., from $U_0 \approx 19$ kV for the $SF_6 : H_2 = 36 : 4.5$ torr mixture). As in [3], we observed a quasi-steady discharge with E/p = 125 V/(cm torr) (Fig. 1a) at low U_0. Upon current passage, the voltage across the plasma decreased as the charge voltage grew (Fig. 1b). In the case of inductive pumping, the residual voltage was relatively low and virtually independent of the charge voltage. The inductive storage substantially raised the breakdown voltage, providing a short current pulse with an amplitude of up to 30 kA to form. In this case, the energy was applied mainly at the trailing edge of the voltage pulse when the electric field intensity is high (Fig. 1c).

Figure 2 shows the dependences of the laser output and efficiency (in terms of the energy deposit) on the charge voltage for inductive and capacitive pumping. In the capacitively pumped $SF_6 : H_2 = 48 : 6$ torr mixture, lasing starts at a charge voltage of ≈26 kV. The maximum efficiency amounts to 11%. The increase in the charge voltage significantly depresses the efficiency with respect to the energy deposit, but the efficiency in terms of the stored energy increases. As was mentioned above, as the charge voltage grows, the residual voltage across the gap and the plasma voltage during the discharge drop. For the inductive pumping of the same mixture, lasing arises at a charge voltage of ≈20 kV. The maximum efficiency in terms of the energy deposit also reaches ≈11% in this case but remains almost unchanged with the charge voltage increasing from 22 to 30 kV.

In the case of the capacitive generator, the maximum efficiency is reached at relatively low current densities and very low charge voltages. The voltage across the gap decreases with increasing current density, which reduces the efficiency in terms of the energy deposit. In the case of pumping by an X-ray-initiated discharge [4], the maximum efficiency in terms of the energy deposit (8%) was also obtained at minimum energy deposits. For each mixture, the appropriate energy deposit was selected by varying the amplitude of the voltage across the discharge gap. The energy deposited into the discharge at the trailing edge of the voltage pulse with high values of the ratio E/p in the laser gap increases substantially if we use inductive pumping. Under these conditions, high-efficiency pumping is also possible at high energy deposits.

The spectral and temporal characteristics of the radiation show that the spectrum of the HF laser in the case of maximum efficiency is much broader than that under conventional pumping conditions [1]. In the spectrum of the high-efficiency laser, the number of rotational lines for each of the vibrational transitions $P(3–2)$, $P(2–1)$, and $P(1–0)$ of the ground electron state increases to 8–11, so that the total number of lines reached 30. In addition, in this case, the powers at the rotational lines of the above vibrational transitions are nearly the same and sequential lasing at several rotational lines of the vibrational transitions is observed. The time–amplitude characteristics of lasing at individual rotational lines indicate that the radiation of $6 \longrightarrow 7$, $7 \longrightarrow 8$, and $8 \longrightarrow 9$ rotational lines of the $P(3–2)$ vibrational transition contributes to the radiation of $7 \longrightarrow 8$, $8 \longrightarrow 9$, and $9 \longrightarrow 10$ rotational lines of the $P(2–1)$ vibrational transition. Taken together, these lines contribute to the radiation of $8 \longrightarrow 9$, $9 \longrightarrow 10$, and

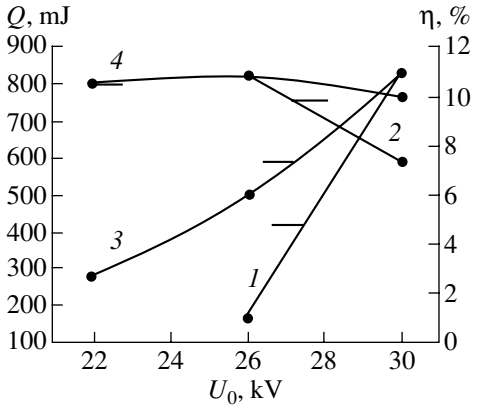


Fig. 2. (*1*, *3*) HF laser output and (*2*, *4*) efficiency in terms of the energy deposit versus the charge voltage for (*1*, *2*) capacitive and (*3*, *4*) inductive pumping. $SF_6 : H_2 = 8 : 1$ mixture, the pressure $p = 0.072$ atm.

$10 \longrightarrow 11$ rotational lines of the $P(1–0)$ vibrational transition. The longest emission pulses are due to the rotational lines of the $P(1–0)$ vibrational transition.

(4) Thus, high efficiency (≈10%) in terms of the energy deposit can be provided by two pumping regimes. In the case of capacitive pumping, the maximum efficiency is attained at the minimum possible discharge-initiating voltage, when the residual voltage after discharge quenching and the plasma voltage are maximal during pumping. However, in this method, the specific energy deposit and output are low. The laser efficiency can be significantly increased using a generator built on a line with distributed parameters whose impedance equals the plasma resistance during pumping.

In the case of inductive pumping, a high efficiency is observed in a wide range of charge voltages and energy deposits. Specifically, the specific laser energy was increased to 4.5 J/l (100 J/(l atm)) at an efficiency ≈10%.

ACKNOWLEDGMENTS

The work was supported by the International Science and Technology Center (project no. 1206).

REFERENCES

1. F. G. Goryunov, K. V. Gurkov, M. I. Lomaev, *et al.*, Kvantovaya Élektron. (Moscow) **21**, 1148 (1994).
2. V. V. Apollonov, S. Yu. Kazantsev, V. F. Oreshkin, and K. N. Firsov, Pis'ma Zh. Tekh. Fiz. **22** (24), 60 (1996) [Tech. Phys. Lett. **22**, 1026 (1996)].
3. E. H. Baksht, A. N. Panchenko, and V. F. Tarasenko, IEEE J. Quantum Electron. **35**, 261 (1999).
4. L. Richeboeuf, S. Pasquiers, F. Doussiet, *et al.*, Appl. Phys. B **68**, 45 (1999).
5. V. P. Borisov, V. V. Burtsev, S. V. Velikanov, *et al.*, Kvantovaya Élektron. (Moscow) **30**, 225 (2000).
6. V. D. Bulaev, V. V. Kulikov, V. N. Petin, and V. I. Yugov, Kvantovaya Élektron. (Moscow) **31**, 218 (2001).
7. V. F. Tarasenko, V. M. Orlovskiĭ, and A. N. Panchenko, Kvantovaya Élektron. (Moscow) **31**, 1035 (2001).
8. V. V. Apollonov, A. A. Belevtsev, S. Yu. Kazantsev, *et al.*, Kvantovaya Élektron. (Moscow) **32**, 95 (2002).
9. E. Kh. Baksht, A. N. Panchenko, and V. F. Tarasenko, Kvantovaya Élektron. (Moscow) **30**, 506 (2000).

Translated by A. Chikishev

Technical Physics, Vol. 48, No. 2, 2003, pp. 270–271. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 139–140.

BRIEF COMMUNICATIONS

Resistance Law for Taylor–Couette Turbulent Flow at Very High Reynolds Numbers

A. M. Balonishnikov

St. Petersburg State University of Engineering and Economics, St. Petersburg, 197002 Russia

e-mail: balonalex@yahoo.co.uk

Received July 1, 2002

Abstract—New expressions for the resistance law and dimensionless moment of force are derived for a Taylor–Couette turbulent flow starting from the generalized model of local balance for the turbulent energy. In the case of extremely high Reynolds numbers, the formulas derived involve a single empirical (Karman) constant.

None of the available turbulence models derived from the Navier–Stokes equations can be considered to be sufficiently accurate. Therefore, comparing semi-empirical turbulence models with experimental data is of great importance. A Taylor–Couette flow between two rotating coaxial cylinders is very convenient for such a comparison. Here, we will consider the case when the inner cylinder rotates and the outer cylinder is at rest.

The basic parameters of such a flow are the Reynolds number $\mathrm{Re} = \Omega a(b-a)/\nu$ and the dimensional moment of forces acting on the inner cylinder $G = T/(\rho\nu^2 L)$, where Ω is the angular velocity of rotation of the inner cylinder, T is the moment of forces, ρ is the density of the liquid, ν is the molecular kinematic viscosity, L is the length of either cylinder, a is the radius of the inner cylinder, and b is the radius of the outer cylinder.

It has been shown [1] that the relationship $G \propto \mathrm{Re}^{\alpha}$ with $\alpha = 5/3$, which was suggested in [2, 3], contradicts experimental data. Agreement was reached by using the Prandtl–Karman resistance law

$$\frac{1}{\sqrt{f}} = c_1 \log(\mathrm{Re}\sqrt{f}) + c_2, \qquad (1)$$

where $f = G/\mathrm{Re}^2$ and the constants c_1 and c_2 are related to the Prandtl–Karman constants.

This law was deduced by joining together the logarithmic profiles of the velocity in the middle of a cylindrical channel. In [4], the relationship $U_f/U \propto \mathrm{Re}^{\xi}$ has been suggested, where U_f is the friction speed on the surface of the inner cylinder, U is the linear velocity of rotation of the inner cylinder ($U = a\Omega$), and $\xi = -0.051$ is the empirical parameter derived from experiments [1]. The value of G was found by fitting data obtained in the same experiments [1]. Eventually, the following expression was derived:

$$G = c_1'\mathrm{Re}^{3/2 + 5\xi/2} + c_2\mathrm{Re}^{2 + 3\xi}, \qquad (2)$$

where $c_1' = 10.5$ and $c_2' = 0.196$.

Now we will show that logarithmic corrections ln(Re) make it impossible to find a power law for the moment of forces. In [5], the Prandtl–Karman law of resistance was deduced from the generalized model of local balance

$$4\pi q u_{\mathrm{in}}^* \kappa^{-1}(q^4-1)^{-1} J_0 + u_{\mathrm{in}}^* q\kappa^{-1}\ln[CU_{\mathrm{in}}^*(b-a)] + u_{\mathrm{put}}^*\kappa^{-1}\ln[Cu_{\mathrm{out}}^*(b-a)/\nu] = V_2 + V_1 q, \qquad (3)$$

where $\mathrm{Re} = V_1(b-a)/\nu$, $q = b/a$, a and b are the radii of the cylinders, u_{in}^* and u_{out}^* are the friction speeds on the inner and outer surfaces of the cylinders, $\kappa = 0.4$ and $C = 9.5$ are the Prandtl–Karman constants for the logarithmic profile of the mean velocity, V_1 and V_2 are the linear velocities of the oppositely rotating cylinders ($V_2 = 0$ for the cylinder at rest), and

$$J_0 = \int_1^q dx\{x\,\mathrm{cosec}[\pi(x^4-1)/(q^4-1)] - (q^4-1)(4\pi)^{-1}(x-1)^{-1} - (q^4-1)(4\pi)^{-1}q^{-2}(q-x)^{-1}\}.$$

Using the well-known relationship for a Taylor–Couette turbulent flow, $u_{\mathrm{out}}^* b = u_{\mathrm{in}}^* a$, and introducing the dimensionless friction speed $z = u_{\mathrm{in}}^*/V_1$, we obtain an algebraic equation for $z = z(\mathrm{Re})$:

$$4q\pi\kappa^{-1}(q^4-1)^{-1}J_0 + q\kappa^{-1}\ln(C\mathrm{Re}z) + \kappa^{-1}q^{-1}\ln(C\mathrm{Re}z/q) = q/z. \qquad (4)$$

1063-7842/03/4802-0270$24.00

If $\mathrm{Re} \longrightarrow \infty$, $z \longrightarrow 0$ and we can derive an asymptotic expression for $z = z(\mathrm{Re})$. First, we omit additive constants. After straightforward rearrangements, we obtain

$$\kappa^{-1}(q + q^{-1})\ln \mathrm{Re} + \kappa^{-1}(q + q^{-1})\ln(z) = q/z. \quad (5)$$

Second, we neglect the second term on the left of (5) to come to

$$z = \frac{\kappa q}{(q + q^{-1})\ln \mathrm{Re}}. \quad (6)$$

It becomes obvious that the second omission is valid, since

$$\lim_{\mathrm{Re} \to \infty} \frac{\ln\ln \mathrm{Re}}{\ln \mathrm{Re}} = 0. \quad (7)$$

The first omission is also valid. Thus, we have

$$z = \frac{\kappa q^2}{(1 + q^2)\ln \mathrm{Re}}. \quad (8)$$

It is known [5] that the coefficient of resistance is given by $c_f = 2\pi z^2 (q-1)^{-2}$ and $G = c_f \mathrm{Re}^2$. Final expressions are those at $\mathrm{Re} \longrightarrow \infty$:

$$c_f = \frac{2\pi\kappa^2 q^4}{(q^2+1)^2 (q-1)^2 (\ln \mathrm{Re})^2}, \quad (9)$$

$$G = \frac{2\pi\kappa^2 q^4}{(q^2+1)^2 (q-1)^2} \left(\frac{\mathrm{Re}}{\ln \mathrm{Re}}\right)^2. \quad (10)$$

These formulas contain the single empirical (Karman) constant $\kappa = 0.4$. This constant and also the logarithmic factor $(\ln \mathrm{Re})^{-2}$ differentiate them from those derived in previous approaches. The table summarizes the values of G_∞ (expression (10)), theoretical values G_{theor} (expression (4)), and associated experimental data G_{expt} [1]. The empirical and theoretical dependences are seen to be in fairly good agreement.

Experimental [1] and theoretical moments of forces G vs. Reynolds number (inner cylinder rotates, outer cylinder is at rest) for the ratio of the radii $q^{-1} = \eta = 0.724$

Re	G_{expt}	G_{theor}	G_∞
10^4	5.0×10^6	5.4×10^6	3.5×10^6
10^5	3.0×10^8	3.3×10^8	2.3×10^8
10^6	1.5×10^{10}	2.2×10^{10}	1.6×10^{10}

Clearly, the value $\mathrm{Re} = 10^6$ is insufficiently high in order for Eq. (3) to be replaced by its approximation (10). However, one should bear in mind that Eq. (10) merely follows from the model of turbulence. It is believed that additional effort should be made to bridge the gap between the direct numerical simulation of the Navier–Stokes equations and cascade models of turbulence used in [1] for comparison with experimental data.

REFERENCES

1. G. S. Lewis and H. L. Swinney, Phys. Rev. E **59**, 5457 (1999).
2. A. Barsilon and J. Brindley, J. Fluid Mech. **143**, 429 (1984).
3. G. P. King, Y. Li, H. L. Swinney, and P. S. Marcus, J. Fluid Mech. **141**, 365 (1984).
4. B. Eckhardt, S. Grossman, and G. Lohse, Eur. Phys. J. B **18**, 541 (2000).
5. A. M. Balonishnikov, Phys. Rev. E **61**, 1390 (2000).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 2, 2003, pp. 272–273. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 2, 2003, pp. 141–142.

BRIEF COMMUNICATIONS

Surface Tension of Pure Aluminum Melt

Kh. Kh. Kalazhokov, Z. Kh. Kalazhokov, and Kh. B. Khokonov

Kabardino-Balkar State University, ul. Chernyshevskogo 173, Nalchik, 360004 Russia

e-mail: ftt@kbsu.ru

Received April 17, 2002

Abstract—The basic reason for the significant spread in the experimental data on the surface tension (σ) of aluminum is shown to be the adsorption of a residual gas in a vacuum chamber. The processing of the kinetic curves $\sigma(\theta)$ by a method proposed in this work yields for aluminum $\sigma = 1070$ mJ/m^2. © *2003 MAIK "Nauka/Interperiodica".*

Published data on the surface tension (ST) of aluminum range from 850 to 1100 mJ/m^2 (Fig. 1) [1, 2]. Values near the upper limit are considered overstated and should be used with caution [2]. Indeed, if σ is determined under conditions when the metal is in thermodynamic equilibrium with its vapor, which is possible only in small sealed chambers disconnected from a pump (as a rule, glass chambers), the resulting value is closer to the lower limit of this interval (Fig. 1, empty circles). On the other hand, if σ is measured with modern setups under conditions of the continuous evacuation of the chamber and detection of surface contamination by an electron spectrometer (Fig. 1, filled circles), the thermodynamic equilibrium of the metal with its vapor breaks and the ST values approach the upper limit of the interval. Thus, the true ST of pure aluminum remains unclear.

Various factors that may influence the ST of metals were the concern of [3]. The key one is the adsorption of the residual gas components on the metal surface or the adsorption relaxation of the as-prepared or cleaned metal surface. The ST variation during surface relaxation under measurement conditions can be expressed by the formula [3]

$$\ln[\sigma(t) - \sigma_r] = \ln[\sigma_0 - \sigma_r] - \frac{t}{\tau_r}, \quad (1)$$

where σ_0 and σ_r are the STs of the metal at the beginning of, and within a sufficiently long time after, the formation (or cleaning) of the surface, respectively, and τ_r has the meaning of the time of surface adsorption relaxation.

Let us apply Eq. (1) to experimental data in [4], where the ST of aluminum was studied as a function of the oxygen content on its surface. To do this, we employ the time dependence of the coverage [5]

$$\theta(t) = \theta_s(1 - \exp[-t/\tau_r]), \quad (2)$$

where θ_s corresponds to the case of equilibrium at a given temperature and pressure.

Assuming that Eqs. (1) and (2) are valid at small θ (or t), expanding Eq. (2) in the Taylor series for small times ($t \ll \tau_r$), and disregarding all the terms except the linear one, we obtain

$$\frac{\theta(t)}{\theta_s} \approx \frac{t}{\tau_r}. \quad (3)$$

Here, θ_s should evidently be replaced by θ_r, which is the coverage at the time moment τ_r. Then, Eqs. (1) and (3) yield

$$\ln(\sigma - \sigma_r) \approx (\sigma_0 - \sigma_r) - \frac{\theta}{\theta_r}. \quad (4)$$

Expression (4) was used to process the experimental dependence $\sigma = \sigma(\theta)$ obtained in [4] for the aluminum surface upon oxygen adsorption. The resulting curve (4) is presented in Fig. 2.

It is seen that the data points can be well approximated by a straight line, thus indicating the validity of expression (4). The initial ST value of aluminum, σ_0, was determined from the intersection of straight line (4) with the vertical axis and turned out to equal 1070 mJ/m^2. This value is in fairly good agreement with the experimental values found by direct measure-

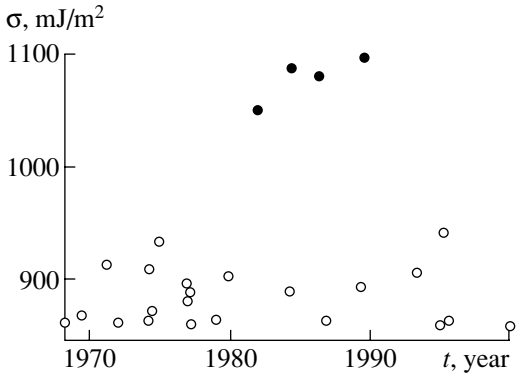


Fig. 1. Surface tension of aluminum vs. the date of measurement [2]. (○) As-prepared and (●) ion-cleaned surface.

1063-7842/03/4802-0272$24.00 © 2003 MAIK "Nauka/Interperiodica"

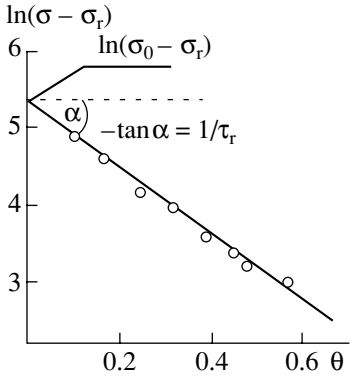


Fig. 2. Determination of σ_0 and θ_s for aluminum from data [4].

ments [4, 6]. When determining σ_0, we found that the ST of aluminum varies during the adsorption relaxation of the surface and, under the conditions in [4], takes the value $\sigma_r = 890$ mJ/m^2 at the time moment $t = \tau_r$. From the slope of straight line (4), the coverage θ_r of the aluminum surface by oxygen was determined to be 0.24. A further increase in θ will evidently lower the ST and lead to better agreement with experimental values of σ measured in sealed devices.

Thus, the ST value approaching 1100 mJ/m^2 [4, 6] should be assigned to the as-prepared or cleaned aluminum surface that has not yet reached equilibrium with a residual gas in the camera. The value of about 860 mJ/m^2 is characteristic of the aluminum surface that is in equilibrium with a residual gas, i.e., is partially covered by gas molecules. The review [2] of methods for determining the ST of aluminum suggests that, with an advance in the experimental instruments and methods, measured ST values increase and approach the upper limit of the interval mentioned above, i.e., come close to 1100 mJ/m^2.

REFERENCES

1. V. I. Nizhenko and L. I. Floka, *Surface Tension of Liquid Metals and Alloys: A Handbook* (Metallurgiya, Moscow, 1981).
2. B. B. Alchagirov, A. M. Chochaeva, Kh. B. Khokonov, *et al.*, Vestn. Kabard.-Balkar. Gos. Univ., Ser. Fiz., No. 4, 4 (2000).
3. Kh. Kh. Kalazhokov and Z. Kh. Kalazhokov, Metally, No. 4, 21 (2000).
4. L. Goumiri and J. C. Jout, Acta Metall. **30**, 1397 (1992).
5. Z. Kh. Kalazhokov, Kh. Kh. Kalazhokov, and Kh. B. Khokonov, *Physics of Interphase Phenomena and Interaction of Particle Fluxes with Solids* (Kabard.-Balkar. Gos. Univ., Nal'chik, 1998), pp. 91–93.
6. A. Pamies, C. Garcia Corbovilla, and E. Louis, J. Mater. Sci. **21**, 2787 (1986).

Translated by A. Sidorova