THEORETICAL AND MATHEMATICAL PHYSICS

Two-Frequency Perturbation of a Smooth Hamiltonian System

V. V. Vecheslavov

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences,
pr. Akademika Lavrent'eva 11, Novosibirsk, 630090 Russia
e-mail: vecheslavov@inp.nsk.su
Received December 30, 2002

Abstract—This work elaborates upon previous studies on the family of smooth continuous and discontinuous two-parameter Hamiltonian systems with a piecewise linear force. For such systems, the Melnikov–Arnold integral is found to be a power and oscillatory function of frequency. In the presence of two primary forcing frequencies, the secondary harmonic with a frequency that is the sum of the primary frequencies may make a major contribution to the formation of a chaotic layer. For the corresponding smooth map, the perturbation parameter ranges where, under strong local chaos, the upper separatrix of fractional resonances is retained while the lower breaks (and vice versa) are determined. It is shown that the zero angle of intersection of the separatrix branches at the central homoclinic point is not a sufficient condition for separatrix retention. Under dynamic conditions, smooth and analytical systems behave in a very different manner. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The occurrence and development of dynamic chaos depend not only on the perturbation magnitude but also on the smoothness of a dynamic system. The smoothness can be conveniently characterized by the rate of decay of Fourier amplitudes. In the analytical case, the amplitudes decay exponentially and there always exists the perturbation threshold ε_{th} above which, $\varepsilon \gtrsim \varepsilon_{th}$, global chaos covering the entire phase space accessible for the system may occur [1–3]. If the potential is a smooth function, its Fourier amplitudes decay as the amplitude number to the power $\beta + 1$, which appreciably changes the character of motion (see, e.g., [4] and references cited there). A time-varying smooth system with one degree of freedom always has a threshold $\varepsilon_{th} > 0$ if $\beta > \beta_{cr} = 3$, where β_{cr} is the critical value of β [4, 5]. The behavior of systems with $\beta < \beta_{cr}$ has remained poorly understood until recently.

It turns out, however, that there have long been available mathematical works [6–8] in which the existence of global invariant curves in smooth systems at $\beta = 2 < \beta_{cr}$ is ascertained. Most comprehensive among them is [7], which concerns symmetric piecewise linear 2D mapping ($\beta = 2 < \beta_{cr}$). In [7], it is rigorously proved that, at certain perturbation parameters, global invariant curves with rational numbers of rotation include the unbroken separatrices of integer and fractional resonances. The separatrices are completely extended in phase and thus present an impenetrable barrier to other trajectories. This excludes the possibility of global action diffusion. Here, most intriguing and unexpected is the fact that the system remains nonintegrable in this

case and the separatrices persist under strong local chaos [9, Fig. 4]. Such behavior contrasts sharply with the situation in typical (i.e., nonintegrable) analytical systems, where, if at least one additional resonance exists, the separatrices of resonances break first, giving rise to chaotic layers [1–3]. The history of this issue and reasons why such an important work has not been properly appreciated are described elsewhere [5, 10].

Later and independently, Ovsyannikov stated the theorem that integer resonance separatrices are retained under symmetric piecewise linear mapping [11]. He strictly defined a set of associated (critical) values of the perturbation parameter and also derived a simple expression for the separatrix in explicit form. The Ovsyannikov theorem has stimulated extensive investigation into piecewise linear and related mappings [5, 9, 10, 12–14] (for a complete statement of the Ovsyannikov theorem, see the appendices to [9, 12]). Note that the authors of [7, 11] had to study only remaining separatrices, since random trajectories resulting when the separatrices split are impossible to treat analytically (these trajectories can be analyzed only in physical or numerical experiments).

Further investigation has shown that each of the global invariant curves discovered in [7] (including the separatrices of fractional and integer resonances) appears at a certain exact value of the perturbation parameter and actually considerably distorts the structure of the phase plane and its finite vicinity. Because of this, a new term, a virtual invariant curve, has been suggested [5]. The presence of virtual curves causes a totally new and very complicated transport process in a

smooth system, so-called fractal diffusion, the study of which has begun in the very recent past [5, 10].

It becomes clear from the above that smooth systems offer many unique properties and dynamic chaos in them originates and develops following a specific (sometimes exotic) scenario.

In this work, as in previous ones [5, 10, 14], we consider a family of smooth systems with two parameters and a piecewise linear force in two versions: with a continuous Hamiltonian and as a map (Section 1). For a continuous system, the Melnikov–Arnold integral is constructed and the frequency dependence of the separatrix map amplitudes is then found using this integral. It is known that harmonics of a separatrix map decay exponentially and monotonically with increasing frequency [1]. In Section 2, we demonstrate that this dependence is radically different for the smooth systems we are considering: it is a power and oscillatory function.

The formation of a chaotic layer subject to an asymmetric two-frequency perturbation is considered. For the analytical case, a similar study has been recently performed in [15–17], where the subject of investigation was the Hamiltonian of a perturbed pendulum:

$$H(x, p, t) = \frac{p^2}{2} + \cos(x)$$
$$+ \varepsilon_1 \cos(x - \Omega_1 t) + \varepsilon_2 \cos(x - \Omega_2 t). \tag{1}$$

Even early numerical experiments showed that the separatrix map spectrum, along with the frequencies Ω_1 and Ω_2, which enter into perturbation (1) in explicit form, also contains combination harmonics $\sim \varepsilon_1 \varepsilon_2$ at the aggregate, $\Delta\Omega_+ = \Omega_1 + \Omega_2$, and difference, $\Delta\Omega_- = \Omega_2 - \Omega_1$, frequencies. Still more surprising is the fact that, under certain conditions, these combination harmonics play a decisive role in the formation of a chaotic layer. An example where the contribution of the secondary harmonic $\Delta\Omega_+ = 3$ to the separatrix map amplitude several hundred times exceeds those from primary harmonics is given in [16]. It appears as though weak primary harmonics generate an intense secondary one and the role of the former in chaos formation is exhausted at this point (this fact has been numerically confirmed in [16]). Taking into consideration the mechanism behind the occurrence and effect of secondary harmonics has made it possible to remove the long known more than twofold discrepancy between the theoretically found (through the Melnikov–Arnold integral) and experimentally measured separatrix map amplitudes for conventional Chirikov mapping [15].

As is shown in Section 3, in the systems under study, the harmonic at the aggregate frequency shows up in full measure, while the harmonic at the difference frequency is absent. This feature is associated with the radical difference in the frequency dependence of the Melnikov–Arnold integral in the analytical and smooth cases. How much the amplitudes and spectral composi-

tions of various parts of the chaotic layer may differ can be inferred from Fig. 2.

In Section 4, mapping is considered again. With the 1 : 3 fractional resonance, it is shown here that there are two ranges of the perturbation parameter K where the upper and lower separatrices behave in a different manner. In the first range, $1/4 \le K \le 1/3$, both separatrices persists for specific values of $K = K_{3, m}$ ($m = 1, 2, \ldots$) [9, Table 1]. With $K < 1/4$, the upper separatrix persists at some values of K, while the lower is retained at others. Such a situation occurs alternately (Fig. 3). The mechanism underlying this intriguing effect is discussed. It is argued that the same behavior is observed for higher order resonances.

In [9, 12, 13], the retention of resonance separatrices was studied by measuring the angle of intersection of branches at the central homoclinic point. It was assumed that the vanishing of this angle proves the retention of the separatrix and absence of the chaotic layer. Such an assumption is mistaken, as demonstrated by an example in Section 5.

1. HAMILTONIAN OF THE PROBLEM

In this work, we elaborate upon the investigation [12, 14] of a continuous system with a Hamiltonian in the form

$$H(x, p, t) = H_0(x, p) + U(x, t),$$
$$H_0(x, p) = \frac{p^2}{2} + \omega_0^2 V(x) \tag{2}$$

and with a two-frequency asymmetric perturbation

$$U(x, t) = \varepsilon_1 \cos(2\pi x - \Omega_1 t) + \varepsilon_2 \cos(2\pi x - \Omega_2 t). \tag{3}$$

The perturbation is assumed to be weak ($|\varepsilon| \ll 1$), and the frequencies are taken to be high ($|\Omega| \gg \omega_0$).

The potential of this system $V(x) = 1/4 - \int f(x)dx$ is generated by an asymmetric piecewise linear force $f(-x) = -f(x)$ with a period of 1:

$$f(x) = \begin{cases} 2x/(1-d), & \text{for} \quad 0 \le x < (1-d)/2 \\ (1-2x)/d, & \text{for} \quad (1-d)/2 \le x \le (1+d)/2 \\ 2(x-1)/(1-d), & \text{for} \quad (1+d) < x < 1. \end{cases} \tag{4}$$

Expression (4) includes the skewness parameter d ($0 < d < 1$), which allows the family of sawteeth to be studied as a whole [14, Fig. 1]. Note that the case of a symmetric force with $d = 1/2$ has been explored most carefully to date. The teeth of the saw $|f(x)| = 1$ are located at the points $x_{-1} = (1 - d)/2$ and $x_{+1} = (1 + d)/2$, where a singularity, a step in the first derivative $f' = df/dx$, is observed:

$$\Delta f'(x_{\mp 1}) = \mp \frac{2}{d(1-d)}. \tag{5}$$

The potential can be represented by the Fourier series [5]

$$V(x) = \frac{1}{4} + \sum_{n \geq 1} \frac{f_n}{2\pi n^{\beta+1}} \cos(2\pi n x),$$

$$f_n = -\frac{2}{\pi^2} \frac{\cos(n\pi)\sin(n\pi d)}{d(1-d)},$$

(6)

where $\beta = 2$ is the index of the system's smoothness.

When unperturbed, system (2) is integrable and describes the only (fundamental) resonance in the vicinity of which initial conditions are defined. Each of the perturbation harmonics is also a resonance. With $\Omega > 0$ or < 0, this resonance is located in the phase plane above or below the fundamental resonance. Therefore, we may call them the upper and lower resonances, respectively.

The motion along the upper unperturbed separatrix ($\varepsilon_1 = \varepsilon_2 = 0$) is described by a function of dimensional time $\psi = 2\omega_0 t$:

$$x_s(\psi_s) = \begin{cases} A_d \exp(\psi_s/\sqrt{2(1-d)}), \\ \quad \text{for } -\infty < \psi_s < -\psi_{s,1} \\ (1 + \sqrt{d}\sin(\psi_s/\sqrt{2d}))/2, \\ \quad \text{for } -\psi_{s,1} \leq \psi_s \leq \psi_{s,1} \\ 1 - A_d \exp(-\psi_s/\sqrt{2(1-d)}), \\ \quad \text{for } \psi_{s,1} < \psi_s < \infty. \end{cases}$$

(7)

The momentum is found by differentiation: $p_s = x = 2\omega_0 dx_s/d\psi_s$. Here, $\psi_{s,1} = \sqrt{2d}\arcsin\sqrt{d}$ and $A_d = (1 - d)\exp(\psi_{s,1}/\sqrt{2(1-d)})/2$ (see formulas (A1.4) and (A1.8) in Appendix 1).

The relative departure from the unperturbed separatrix in terms of relative energy is designated as $\omega = H_0/H_{0,s} - 1$, where $H_{0,s} = \omega_0^2/4$ is the Hamiltonian on the separatrix. The motion period T_0 near the separatrix is calculated by the formula

$$T_0(w) = 2T_{s,1} + \frac{1}{\omega_2}\ln\frac{4\sqrt{1-d}\cos(\omega_1 T_{s,1})}{|w|},$$

(8)

where $\omega_1 = \omega_0\sqrt{2/d}$, $\omega_2 = \omega_0\sqrt{2/(1-d)}$, and $T_{s,1} = \arcsin\sqrt{d}/\omega_1$ (see Appendix 1).

When the dimensions of the chaotic layer are determined numerically, it is convenient to use the relationship between the relative energy w and motion period T_0 that is inverse to (8):

$$w(T_0) = 4\sqrt{(1-d)}\cos(\omega_1 T_{s,1})\exp(-\omega_2(T_0 - 2T_{s,1})). \quad (9)$$

Along with the continuous system given by (2) and (3), we will also address the discrete map with the same potential:

$$\bar{p} = p + Kf(x), \quad \bar{x} = x + \bar{p}\,\mathrm{mod}1, \quad (10)$$

where $K = \omega_0^2$. This map has been much studied to date, and its dynamic behavior appears very unusual [5, 9, 10, 12–14].

As is well known, map (10) can be represented as a continuous system with the Hamiltonian explicitly depending on time and experiencing a perturbation in the form of kicks [5]:

$$H(x, p, t) = \frac{p^2}{2} + KV(x)\delta_1(t), \quad (11)$$

where $\delta_1(t) = 1 + 2\sum_{m \geq 1}\cos(2\pi m t)$ is the delta function of period 1.

The period of force (4) can be subdivided into two intervals where the derivative f' is negative (elliptic interval) and positive (hyperbolic interval). At a fixed point, the matrix of the linearized mapping is determined from the formulas

$$\mathscr{A} = \begin{pmatrix} 1+\Delta & 1 \\ \Delta & 1 \end{pmatrix}, \quad \mathscr{B} = \begin{pmatrix} 1-\Delta & 1 \\ -\Delta & 1 \end{pmatrix}, \quad (12)$$

where $\Delta = 2K/(1 - d)$. The matrices $\mathscr{A}$ and $\mathscr{B}$ correspond to the hyperbolic and elliptic intervals, respectively.

2. FREQUENCY DEPENDENCE OF THE MELNIKOV–ARNOLD INTEGRAL

In the general case, perturbation (3) breaks the fundamental resonance separatrix with the formation of a chaotic layer, which can be subdivided into three regions: (i) upper region where the phase rotates at $p > 0$, (ii) intermediate region where the phase oscillates, and (iii) lower region where the phase rotates at $p < 0$. Under asymmetric perturbation, the dimensions of these regions may differ substantially, since the upper region forms largely as a result of upper resonances; the lower region, as a result of lower resonances; and the intermediate region, as a result of both (Fig. 2). To be specific, we will study the upper region of the chaotic layer.

Consider an upper perturbation harmonic of type $\varepsilon\cos(2\pi m x - \tau - \tau_0)$, where $\tau = \Omega t$ and m is an integer factor introduced for generality. We will seek a change in the unperturbed energy H_0 due to this perturbation over the oscillation or rotation half-period, following

the technique described in [1]:

$$\Delta H_0 = -\varepsilon \int_{-\infty}^{\infty} p(t) \frac{\partial U}{\partial x} dt$$

$$= 2\pi m\varepsilon \int_{-\infty}^{\infty} p(t)\sin(2\pi mx - \tau - \tau_0)dt = 2\pi m\varepsilon \sin\tau_0 W_{MA},$$

where W_{MA} is the Melnikov–Arnold integral

$$W_{MA} = -\int_{-\infty}^{\infty} p_s(t)\cos[2\pi mx_s(t) - \Omega t]dt. \qquad (13)$$

In the expression for the energy change, only the even function in the expansion of $\sin(2\pi mx - \tau - \tau_0)$ is included and the system is assumed to move near the unperturbed separatrix.

In going to the dimensionless time $\psi = 2\omega_0 t$ and calculating expression (13) with the help of (7), we find

$$W_{MA}(\lambda > 0) = \frac{1}{\sqrt{2}} \int_0^{\psi_{s,1}} \cos\left(\frac{\psi}{\sqrt{2d}}\right)$$

$$\times \cos\left[\pi m\sqrt{d}\sin\left(\frac{\psi}{\sqrt{2d}}\right) - \lambda\psi\right]d\psi$$

$$- A_d\sqrt{\frac{2}{1-d}} \int_{\psi_{s,1}}^{\infty} \exp(-\psi/\sqrt{2(1-d)}) \qquad (14)$$

$$\times \cos(2\pi mA_d\exp(-\psi/\sqrt{2(1-d)}) + \lambda\psi)d\psi,$$

where $\lambda = \Omega/2\omega_0$ is the adiabaticity parameter [1].

It is easy to check that, if a perturbation is a lower harmonic of type $\cos(2\pi mx + \tau + \tau_0)$, where $\tau = \Omega t$, (14) must be replaced by

$$W_{MA}(\lambda < 0) = -\frac{1}{\sqrt{2}} \int_0^{\psi_{s,1}} \cos\left(\frac{\psi}{\sqrt{2d}}\right)$$

$$\times \cos\left[\pi m\sqrt{d}\sin\left(\frac{\psi}{\sqrt{2d}}\right) + |\lambda|\psi\right]d\psi$$

$$+ A_d\sqrt{\frac{2}{1-d}} \int_{\psi_{s,1}}^{\infty} \exp(-\psi/\sqrt{2(1-d)}) \qquad (15)$$

$$\times \cos(2\pi mA_d\exp(-\psi/\sqrt{2(1-d)}) - |\lambda|\psi)d\psi.$$

It should be noted that, for $m = 1$ and $d = 1/2$, the half-sum of (14) and (15) exactly coincides with formula (13) from [12], where a symmetric one-frequency perturbation of our system is considered.

The separatrix map amplitude is completely defined by the properties and behavior of the Melnikov–Arnold integral, since it is proportional to this integral:

$$W = \pm\max|\bar{w} - w| = \pm\frac{\Delta H}{H_{0,s}} = \frac{8\pi m\varepsilon}{\omega_0^2}W_{MA}. \qquad (16)$$

The first question to be answered is how W_{MA} varies with λ. Appendix 2 gives asymptotic ($\lambda \longrightarrow \infty$) estimates of the Melnikov–Arnold integral, which lead us to the following conclusions for the system given by (2) and (3). From formulas (A2.2) and (A2.3) in Appendix 2, it follows that W_{MA} is a periodic function of λ with a period

$$T_\lambda = \frac{2\pi}{\psi_{s,1}} = \frac{\pi}{\arcsin\sqrt{d}}\sqrt{\frac{2}{d}}, \qquad (17)$$

and an amplitude $W_{MA,max}$ varying inversely proportionally to λ^3. The value of $W_{MA,max}$ is the same for the upper and lower harmonics:

$$W_{MA,max}(\lambda) \approx \frac{|\lambda|^{-3}}{4d}\sqrt{\frac{1-d}{2}}. \qquad (18)$$

Nevertheless, numerical calculations show that the contributions from the upper and lower harmonic are different. To reveal this circumstance, we had to take into account terms of the order λ^{-4}. This allowed us to find a small difference between the upper and lower harmonics (formulas (A2.4) and (A2.5) in Appendix 2). Eventually, we get

$$W_{MA}(\lambda) \approx (-1)^m \frac{|\lambda|^{-3}}{4d}\sqrt{\frac{1-d}{2}}$$

$$\times \left[1 \pm \pi|\lambda|^{-1}\sqrt{\frac{1-d}{2}}\right]\sin(\pi md \mp |\lambda|\psi_{s,1}), \qquad (19)$$

where the upper signs refer to the upper harmonic.

Figure 1 shows three $W_{MA}|\lambda|^3$ dependences of the reduced Melnikov–Arnold integral for the symmetric saw, $d = 1/2$. With $d \neq 1/2$, the dependences remain basically the same but the harmonics shift in phase. It is seen that formulas (19) are in qualitative agreement with these dependences while they need some numerical correction. Note that the zeros of the integral just indicate that the fundamental resonance separatrix remains unsplit [12].

The above results differ drastically from the case of analytical potential (1), where W_{MA} is a function of λ that is monotonic and decreasing everywhere. Moreover, with frequencies that are equal in magnitude, the contribution of the lower harmonic to the upper region of the chaotic layer is $\exp(-\pi|\lambda|)$ times smaller than that of the upper harmonic [1]. In the system given by (2) and (3), the situation is reversed: as the frequency grows, the contributions of the upper and lower harmonics to the separatrix map approach each other. Such a radical difference in the behavior of the two systems

is probably associated with the fact that the positions of singularities in the Melnikov–Arnold integral greatly differ: for smooth potential (2), they are on the real time axis, while for analytical potential (1), they are in the complex plane.

3. SECONDARY HARMONICS IN THE SEPARATRIX MAP

The separatrix map, introduced for the first time in [18], approximates the dynamic behavior of a system at instants when it passes stable equilibrium positions. Elements of the map that are responsible for different parts of the chaotic layer differ in amplitude and also in spectral composition (Fig. 2). Therein lies the qualitative difference between an asymmetric multiple-frequency perturbation and the well-studied one-frequency symmetric perturbation [1].

As was noted in the Introduction, the separatrix map generally includes, along with harmonics at primary frequencies (these harmonics enter Hamiltonian (2) in explicit form), harmonics at combined frequencies (secondary harmonics). Because of this, the separatrix map for the system given by (2) and (3) should be written in the form [17]

$$\overline{w} = w + \sum_{k=1}^{K_u} W_{u,k}(\Omega_{u,k} t_0), \quad \text{if} \quad p > 0, \tag{20}$$

$$\overline{w} = w + \sum_{k=1}^{K_l} W_{l,k}(\Omega_{l,k} t_0), \quad \text{if} \quad p < 0, \tag{21}$$

$$\overline{t_0} = t_0 + T_0(\overline{w}). \tag{22}$$

Here, the subscript u marks resonances contributing significantly to the formation of the upper region of the chaotic layer and the subscript l marks those resonances making a considerable contribution to the formation of the lower region.

Instants t_0 when the system passes stable equilibrium position $x = 0.5 \mod(1)$ are counted on a continuous time scale. If all the frequencies are multiples of some reference frequency Ω_0, expression (22) may be recast as

$$\begin{aligned} \overline{\phi_0} &= \phi_0 + \Omega_0 T_0(\overline{w}), \\ \phi_0 &= \Omega_0 t_0 \mod(2\pi), \end{aligned} \tag{23}$$

although such a conversion is not obligatory in actual practice.

Formulas (14)–(16) from the previous section make it possible to find the amplitude of any harmonic in the separatrix map if its perturbation amplitude ε is known. Primary harmonics pose no problem, since they are explicitly defined by (3). To find the perturbation amplitudes for secondary harmonics, we introduce, following [15], new variables $y(t) = x(t) - x_s(t)$ and $u(t) =$
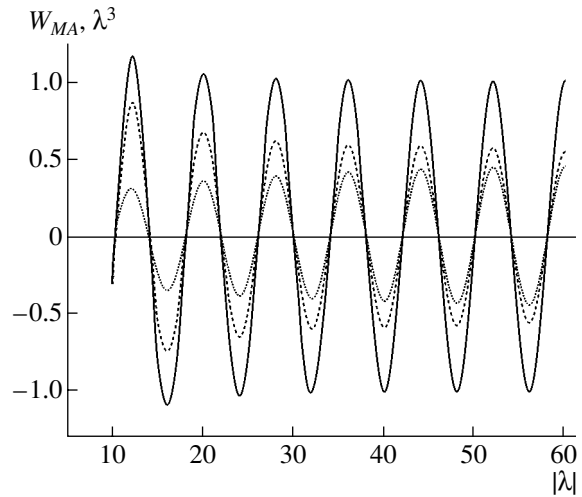

W_{MA}, λ^3

Fig. 1. Melnikov–Arnold integral vs. adiabaticity parameter $\lambda = \Omega/2\omega_0$. The upper curve reflects the combined action of the upper and lower harmonics of the symmetric perturbation; the intermediate curve, the action of the upper harmonic alone; and the lower curve, the action of only the lower harmonic.

$p(t) - p_s(t)$ and then, using a generating function of type $F_2(u, x, t) = [p_s(t) + u][x - x_s(t)]$, pass from (2) to a new Hamiltonian $H(y, u, t)$, which describes motion near the separatrix. Since the perturbation is weak, we expand the potentials $V(x_s + y)$ into the Taylor series up to second order and make approximate substitutions $\sin(2\pi y) \longrightarrow 2\pi y$ and $\cos(2\pi y) \longrightarrow [1 - (2\pi y)^2/2]$ to obtain an equation of motion in the form

$$\begin{aligned} \frac{d^2 y}{dt^2} = y\left[-\omega_0^2\left(\frac{d^2 V}{dx_s^2}\right) + 4\pi^2 \sum_{n=1}^{2} \varepsilon_n \cos(2\pi x_s - \Omega_n t) \right] \\ + 2\pi \sum_{n=1}^{2} \varepsilon_n \sin(2\pi x_s - \Omega_n t). \end{aligned} \tag{24}$$

We will restrict our analysis to a forced (vanishing at $\varepsilon \longrightarrow 0$) solution to (24), which can be derived by the method of successive approximations. If, as in [15],

$$\begin{aligned} y_\varepsilon^{(1)} &= -\sum_{n=1}^{2} \frac{2\pi\varepsilon_n}{(2\pi p_s - \Omega_n)^2} \sin(2\pi x_s - \Omega_n t) \\ &\approx -\sum_{n=1}^{2} \frac{2\pi\varepsilon_n}{\Omega_n^2} \sin(2\pi x_s - \Omega_n t) \end{aligned}$$

is taken as a first approximation (the approximate equality assumed that $\Omega \gg p_{s,\max}$), terms of type $\varepsilon_n \sin(2\pi x_s - \Omega_n t)$ in (24) vanish but new ones appear. The latter can be eliminated by applying the second
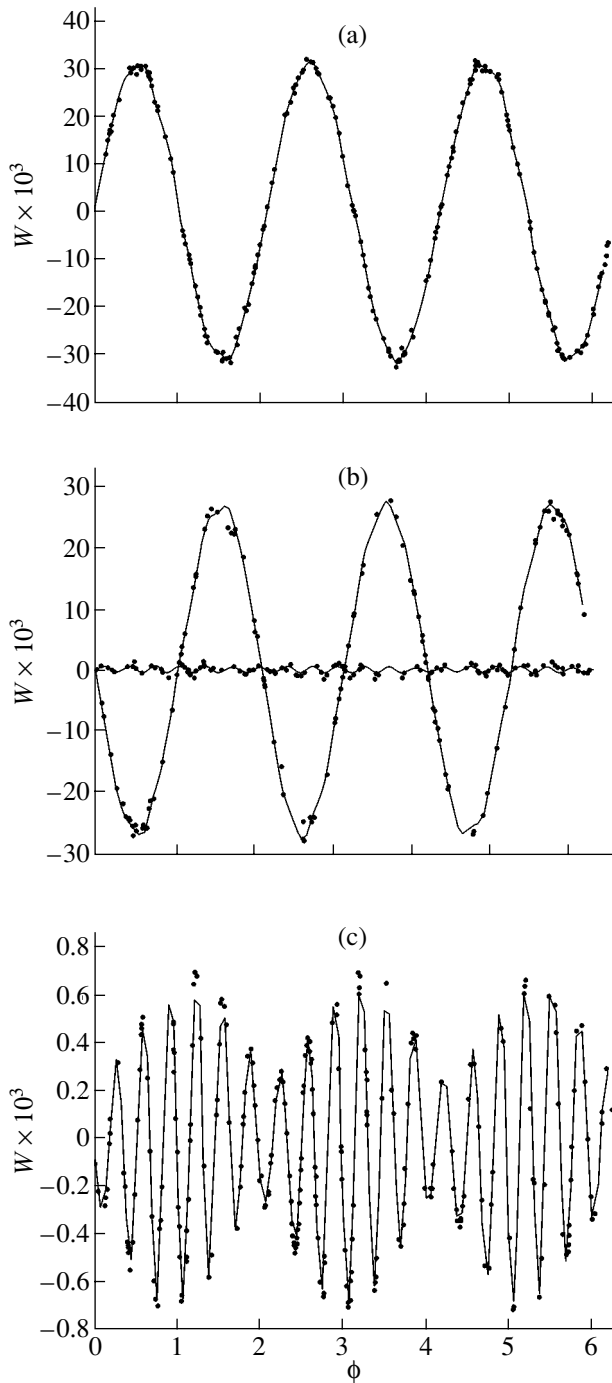
Fig. 2. Separatrix map of system (2)–(3) with asymmetric perturbation (27). The dots were obtained by numerical calculation, and the continuous curves are the least-squares fits. (a) The upper part of the chaotic layer (phase rotation with $p > 0$). The amplitudes of the primary harmonics are $W(23) \approx 6.9 \times 10^{-4}$ and $W(-19) \approx -4.9 \times 10^{-4}$; the amplitude of the secondary harmonic at the aggregate frequency is $W(3) \approx 3.03 \times 10^{-2}$. (b) The middle part of the chaotic layer (phase oscillation). Two periodic curves reflect the action of the upper and lower resonances. (c) The lower part of the chaotic layer (phase rotation with $p < 0$). The amplitudes of the primary harmonics are $W(23) \approx 2.1 \times 10^{-4}$ and $W(-19) \approx -4.8 \times 10^{-4}$; the amplitude of the secondary harmonic is $W(3) \approx 5.6 \times 10^{-5}$.

approximation

$$y_\varepsilon^{(2)} \approx -\sum_{n=1}^{2} \frac{2\pi\varepsilon_n}{\Omega_n^2} \sin(2\pi x_s - \Omega_n t)$$

$$-\frac{2\pi^2\varepsilon_1\varepsilon_2}{(\Omega_2 - \Omega_1)^2}\left[\frac{1}{\Omega_2^2} - \frac{1}{\Omega_1^2}\right]\sin(\Omega_2 - \Omega_1)t + \dots \tag{25}$$

etc. [15].

Turning back to system (2)–(3) and putting $x \approx x_s + y_\varepsilon^{(2)}$, $\cos 2\pi y_\varepsilon^{(2)} \approx 1$, and $\sin 2\pi y_\varepsilon^{(2)} \approx 2\pi y_\varepsilon^{(2)}$, we make sure that secondary perturbation harmonics ($\sim\varepsilon_1\varepsilon_2$) at the aggregate and difference frequencies arise. However, the mechanism of their penetration into the separatrix map is essentially different. The harmonic at the aggregate frequency forms because of the interaction of the first term in (25) with primary harmonics, and, since this exactly coincides with the analytical case, equalities (7) in [16] for it hold (these equations should only be corrected in view of another phase normalization):

$$\varepsilon_+\cos(4\pi x - \Delta\Omega_+ t),$$

$$\varepsilon_+ = -2\pi^2\varepsilon_1\varepsilon_2\left[\frac{1}{\Omega_1^2} + \frac{1}{\Omega_2^2}\right], \tag{26}$$

where $\Delta\Omega_+ = \Omega_1 + \Omega_2$.

The harmonic at the difference frequency appears in another way: it results from the interaction of the second term in (25) with the second derivative d^2V/dx^2 of the potential in (24), which has rich spectrum (6) (unlike (1), where $V(x) = \cos x$). Here, an infinite number of harmonics, rather than two as in [16], appear, and it is impossible to separate out the most significant one.

The effect of occurrence of the secondary harmonic at the aggregate frequency will be demonstrated with system (2)–(3) for

$$\omega_0^2 = 0.09, \quad \varepsilon_1 = \varepsilon_2 = 0.05,$$

$$\Omega_1 = 22.0, \quad \Omega_2 = -19.0, \quad \Omega_0 = 1.0. \tag{27}$$

Let us construct separatrix maps for all three parts of the layer. First, we will briefly recall this procedure (for details, see [15]). On the line $x = 0.5$, the central homoclinic point $\mathcal{P}_{fb}$ is found with a high accuracy (this point is the boundary between phase rotation and phase oscillation). The nonzero angle of intersection of the separatrix branches at this point indicates the presence of the chaotic layer (the reverse is untrue, see Section 5). A narrow interval $\mathcal{P}_{fb} + \delta\mathcal{P}$ is selected on the line $x = 0.5$ near this point within the considered part of the layer, and a random trajectory is injected from this interval. This trajectory either executes periodic motion with a desired number of periods (time intervals T_0 between sequential passages through the phase $x = 0.5$) or is interrupted because of the transition to another part of the layer. In both cases, a new random trajectory is injected from the same interval until a desired number

N_p of periods is reached. For each of the periods, the mean energy $\bar{w}$ is calculated by formula (9). Finding the difference $\delta w = \bar{w} - w$ for each pair of adjacent periods and assigning this change to a time instant t_0 common to this pair, one can construct the separatrix map $(\delta w)_k$, $t_{0,k}$, where $k = 1, 2, \ldots, N_p - 1$, on a continuous time scale.

The points in Fig. 2 were obtained precisely in this way. The continuous lines are least-squares fits by which the amplitudes W of individual harmonics were determined (below and in the caption to Fig. 2, the parenthesized figures are frequencies). The measured contribution $W(3) \approx 3.03 \times 10^{-2}$ of the secondary harmonic to the formation of the upper part of the chaotic layer is more than forty times that of the primary harmonics. The theoretical value W_t of the amplitude $W_t(3)$ of the separatrix map for this frequency, which follows from (14), (16), and (26), equals $W_t(3) \approx 3.37 \times 10^{-2}$ ($\varepsilon_+ \approx 2.39 \times 10^{-4}$, $W_{MA} \approx 0.505$).

In practice, the widths of the parts of the layer (in terms of the relative energy w), rather than the separatrix map amplitudes, are of greater importance. The widths can be found by applying iterations to the separatrix map and by searching for the minimal motion period $T_{0,\min}$ (the width is then determined from formula (9)). When applied to the top part of the layer in case (27), both approaches give close values: $w_{st} \approx 0.19$ (former) and $w_{st} \approx 0.32$ (latter). For the bottom part of the layer, $w_{sb} \approx 0.016$ and $w_{sd} \approx 0.022$, respectively. Note that the former approach demands much less (several hundreds of times) computation time, so that efforts spent on the construction of the separatrix map are justified.

4. ASYMMETRY IN SPLITTING THE UPPER AND LOWER SEPARATRICES

In this section, we again consider map (10) with $d = 1/2$ and discuss the behavior of the separatrices of fractional resonances. First we will recall the structure of a separatrix using a pendulum as an example [1].

The curve has a saddle, a fixed point, which must be considered as an independent trajectory (an undisturbed pendulum may stay at this point infinitely long). Two more trajectories (separatrices) leave the saddle in opposite directions and then asymptotically approach it. Both of them are the boundary between phase rotation (outside a resonance) and phase oscillation (inside a resonance). Near the saddle in the phase plane, a specific cross with two incoming and two outgoing trajectories forms [1, Fig. 2.1]. Two fundamental resonance separatrices with $p_s = p_{st} > 0$ and $p_s = p_{sb} < 0$ can be conveniently called the upper and lower separatrices, respectively. It is important that both unperturbed separatrices actually consist of two trajectories, which are spatially coincident for time scales directed back and forth, respectively. A perturbation splits either of the
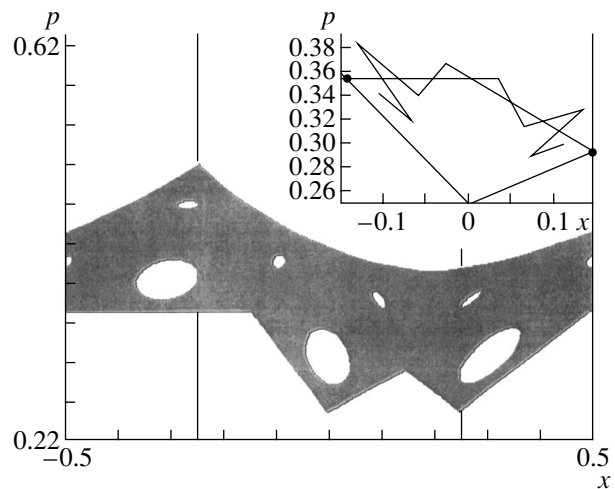

Fig. 3. Map (10) with the parameters $K = 0.1035533905931\ldots$ and $d = 1/2$. The dark region shows the only random trajectory bounded below by the lower separatrix of the 1 : 3 resonance (the number of iterations is 5×10^9). The inset shows the branches (filaments) detached from the upper separatrix and the unbroken lower separatrix of this resonance between two neighboring saddles.

separatrices into two branches (filaments), which are no longer coincident but intersect at homoclinic points. The nonzero angle of intersection of the branches at any of these points (usually, the central homoclinic point is considered; in (10), it corresponds to $x_0 = 0.5$ [9]) is a reliable indication that a separatrix has broken and a random trajectory has taken its place (the reverse is untrue, see Section 5).

As follows from the above, the upper and lower separatrices break under the action of the upper and lower sets of resonances, respectively, and these sets are generally different. As a consequence, the upper and lower parts of the chaotic layer may be dissimilar (Fig. 2). In system (10), a unique effect is observed [7]: the separatrices of both fractional and integer resonances persist under the condition of strong local chaos for specific values of the perturbation parameter K. We will show that this phenomenon for fractional resonances exhibits curious features.

The object of consideration is the 1 : 3 fractional resonance. We will rely upon data obtained in [7] (for extra details, see also [10]). The value $K = 1/3$ is the greatest critical number at which the separatrices of all system resonances (and not only of the 1 : 3 resonance) are retained. These (and only these!) separatrices, which are nontransparent to other trajectories, fill the unit square of the phase plane, the entire area occupied by this mosaic being exactly equal to the area of the square [7; 9, Fig. 6]. As K gradually decreases to $K \geq 1/4$, the system passes through a set of specific values $K_{3,m}$ ($m = 1, 2, \ldots$,) at each of which both (upper and lower) separatrices of the 1 : 3 resonance persist (the initial ten values $K_{3,m}$ are listed in [9, Table 1]). At first glance, such a situation is puzzling. The set of resonances
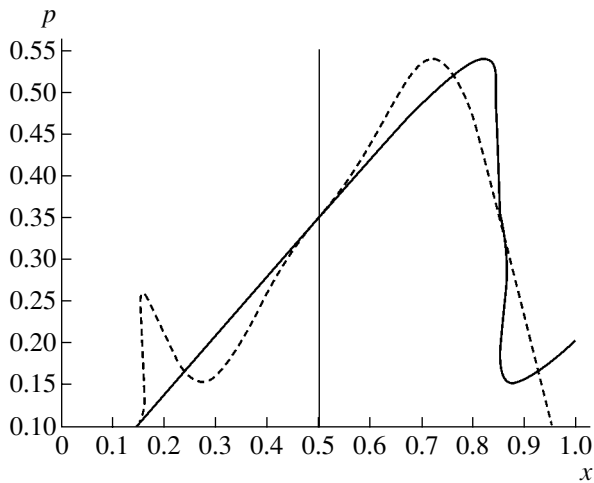
Fig. 4. Map (28) with the parameters $C = 0.32395435\ldots$ and $a = 0.1$. The detached branches of the upper separatrix are shown. The angle at the central homoclinic point equals zero. The continuous and dashed lines correspond to time running forward and backward, respectively.

above and below the resonance considered is by no means symmetric. The nearest strongest resonances above and below are the integer and fractional resonances, respectively. Yet the retention conditions for both separatrices coincide exactly. It would be interesting to understand the reason for this phenomenon.

The set $K_{3,m}$ has a condensation point at $K = 1/4$, where the phase volume of the 1 : 3 resonance goes to zero and its upper and lower separatrices collapse, turning into a broken line. The resonance disappears, leaving behind it a global invariant curve. It was shown [7, Fig. 1] that this curve is a periodic orbit each point of which maps onto itself after three iterations. It is important that all three saddle points of the 1 : 3 resonance are on the hyperbolic segment, as follows from our investigation in the perturbation parameter range $1/4 < K \le 1/3$. With $K < 1/4$, the resonance 1 : 3 arises again but now the upper and lower separatrices behave in a vastly different manner: they persist alternately at various values of K and not simultaneously. This is illustrated in Fig. 3, which is plotted for the first critical value of K, $K = 0.1035533905931$, following $K = 1/4$. At this value, the lower separatrix persists. It should also be noted that, for $K \lesssim 1/4$, one of the saddle points passes from the hyperbolic to elliptic section.

Preliminary numerical experiments show that, as K decreases further, the value at which two saddles pass to the elliptic section will be found, etc. In addition, it turned out that the same behavior is also typical of high order fractional resonances. In general, the situation appears as follows. For each resonance of order $Q \ge 3$, one can find a $K(Q)$ parameter value nearest to $K = 1/3$ such that this resonance degenerates into a periodic orbit (for several Q, these values are listed in [7]). Within this range, all saddles of the resonance fall into the hyperbolic section. Below this range, at least one of

the saddles passes to the elliptic section. As the perturbation parameter smoothly decreases further, the number of saddle points in the elliptic section will grow.

The migration of saddle points throws light on the situation. As is known, the upper and lower separatrices of resonances bypass their saddles in opposing directions. It was noted above that, early in the formation of resonant structures at $K = 1/3$ and in some interval below this value, all saddles of a resonance Q fall into the hyperbolic section and the matrix of a periodic unstable orbit is merely $\mathscr{A}^Q$ (see (12)) irrespective of the by-pass order. The retention conditions for the upper and lower separatrices are exactly the same and occur at the same critical value of the perturbation parameter K. The situation changes radically after the resonance has degenerated into a periodical orbit, since one saddle jumps to the elliptic section and is now described by the matrix $\mathscr{B}$. It is here that asymmetry in the behavior of the separatrices arises, since one of them passes from the hyperbolic to the elliptic section (the matrix $\mathscr{B}\mathscr{A}$), while the other does the opposite (the matrix $\mathscr{A}\mathscr{B} \ne \mathscr{B}\mathscr{A}$). Because of this, the upper and lower separatrices persist alternately at different values of K. The question arises as to whether the separatrices will persist simultaneously if all saddles pass to the elliptic section? The curious effect described above is another specific feature of smooth systems.

5. INSTEAD OF CONCLUSIONS

The objects of consideration in this work were largely nonlinear resonances and their separatrices. It was noted in the introduction that the angle of intersection of the separatrix branches at the central homoclinic point is of key importance in such studies [9, 12, 13]. This quantity is one of a few chaos attributes that can be measured with any desired accuracy. If this angle is other than zero, the separatrix is split and chaos occurs. The major practical issue is whether the equality of this angle to zero is a reliable indication that the separatrix persists. Up to now, this has been the case and such an assertion has even appeared in several of our previous works. However, special research has demonstrated that this statement is untrue. An appreciable number of systems, both continuous and discrete (maps), where the vanishing of this angle does not mean the retention of the separatrix have been found. Let us illustrate this with a complicated standard mapping of type

$$\bar{p} = p + C[\sin(2\pi x) - a\sin(6\pi x)], \quad \bar{x} = x + p \quad (28)$$

with the coefficients $C = 0.32395435\ldots$ and $a = 0.1$. The branches of the upper separatrix are shown in Fig. 4 (the central homoclinic point is on the line $x = 0.5$). The point of contact here corresponds to an inflection point. It seems that only the complete pattern of the separatrix behavior constructed for time running back and forth in the interval between two neighboring

homoclinic points might tackle the question of whether chaos is present or not.

From the above, it follows that the dynamic behavior of smooth and analytical systems differs greatly in terms of almost all the aspects considered. In our opinion, this is a compelling argument in favor of further research into smooth systems.

ACKNOWLEDGMENTS

APPENDIX 1

MOTION ALONG AND NEAR THE UNPERTURBED SEPARATRIX

The Hamiltonian on the unperturbed separatrix equals $H_{0,s} = \omega_0^2/4$. From this expression, we find its momentum as a function of the coordinate:

$$\frac{p_s}{\omega_0} = \begin{cases} x_s\sqrt{2/(1-d)}, & \text{for} \quad 0 < x_s < (1-d)/2 \\ (1-(2x_s-1)^2/d)^{1/2}/\sqrt{2}, \\ \text{for} \quad (1-d)/2 < x_s < (1+d)/2 \\ (1-x_s)\sqrt{2/(1-d)}, & \text{for} \quad (1+d) < x_s < 1. \end{cases} \quad (A1.1)$$

Let dimensionless time $\psi = 2\omega_0 t$ be counted from the point of stable equilibrium $x = x_0 = 0.5$. In the (first) interval $x_0 \le x_1 = (1+d)/2$, the equation of motion is written as $\ddot{x} = -\partial H_0/\partial x = \omega_0^2(1-2x)/d$ and has the solution

$$x(t) = \frac{1}{2} + \frac{p_0}{\omega_1}\sin\omega_1 t, \quad \omega_1 = \omega_0\sqrt{\frac{2}{d}}, \quad (A1.2)$$

where p_0 is the momentum at the initial point $x_0 = 0.5$.

For the separatrix corresponding to the second row in (A1.1), we have at this point $p_{0,s}/\omega_0 = 1/\sqrt{2}$; therefore,

$$x_s(\psi_s) = \frac{1}{2}\left[1 + \sqrt{d}\sin\frac{\psi_s}{\sqrt{2d}}\right],$$
$$p_s(\psi_s) = \frac{\omega_0}{\sqrt{2}}\cos\frac{\psi_s}{\sqrt{2d}}. \quad (A1.3)$$

The time $\psi_{s,1}$ of motion along the separatrix throughout the first interval is found with (A1.3) from the equality

$$\frac{1+d}{2} = \frac{1}{2}\left[1 + \sqrt{d}\sin\frac{\psi_{s,1}}{\sqrt{2d}}\right].$$

At the end of the first interval, we have

$$T_{s,1} = \arcsin\sqrt{d}/\omega_1, \quad \psi_{s,1} = 2\omega_0 T_{s,1} = \sqrt{2d}\arcsin\sqrt{d},$$
$$x_{s,1} = \frac{1+d}{2}, \quad p_{s,1} = \omega_0\sqrt{\frac{1-d}{2}}, \quad (A1.4)$$

where $T_{s,1}$ is dimension (normal) time.

In the second interval $(1+d)/2 < x < 1$, motion is described by the equation $\ddot{x} = 2\omega_0^2(x-1)/(1-d)$, which has the solution

$$x(t) = 1 - \frac{1-d}{2}\cosh\omega_2(t-T_1) + \frac{p_1}{\omega_2}\sinh\omega_2(t-T_1),$$
$$\omega_2 = \omega_0\sqrt{\frac{2}{1-d}}, \quad (A1.5)$$

where T_1 and p_1 are the time and momentum at the end of the first interval.

For motion along the separatrix according to equalities (A1.4), we have

$$x_s(t) = 1 - \frac{1-d}{2}[\cosh\omega_2(t-T_{s,1}) - \sinh\omega_2(t-T_{s,1})]$$
$$= 1 - \frac{1-d}{2}\exp(-\omega_2(t-T_{s,1})). \quad (A1.6)$$

In terms of dimensionless time ψ, motion along the separatrix in the second interval is given by

$$x_s(\psi_s) = 1 - A_d\exp(-\psi_s/\sqrt{2(1-d)}),$$
$$p_s(\psi_s) = A_d\sqrt{\frac{2}{1-d}}\exp(-\psi_s/\sqrt{2(1-d)}), \quad (A1.7)$$

where

$$A_d = \frac{(1-d)}{2}\exp(\psi_{s,1}/\sqrt{2(1-d)}) \quad (A1.8)$$

and $\psi_{s,1}$ is given by (A1.4).

Taken together, the above formulas describe motion along the entire separatrix (see formulas (7) in Section 1).

To determine the period of motion near the separatrix, let us consider a trajectory starting from the initial point $x_0 = 0.5$ with a momentum p_0 slightly deviating from the momentum on the separatrix: $p_0 = p_{s,0} + \delta p_0$. The difference in momentum changes the time $T_{s,1}$ of passing the first interval by δT_1 and the momentum $p_{s,1}$ at the end of the interval by δp_1. The new time can be found from (A1.2):

$$\frac{1+d}{2} = \frac{1}{2} + \frac{p_{s,0} + \delta p_0}{\omega_1}\sin[\omega_1(T_{s,1} - \delta T_1)]$$
$$\longrightarrow \omega_1\delta T_1 = \frac{\delta T_1}{p_{s,0}}\tan(\omega_1 T_{s,1});$$

the new momentum, from the relationship

$$p_1 = p_{s,1} + \delta p_1$$
$$= (p_{s,0} + \delta p_0)\cos[\omega_1(T_{s,1} - \delta T_1)]$$
$$\longrightarrow \delta p_1 = \frac{\delta p_0}{\cos(\omega_1 T_{s,1})}.$$

The latter quantity can be expressed through the relative change in the energy $w = 4H_0/\omega_0^2 - 1$.

Since $V(x_0 = 0.5) = 0$, we have

$$w = \frac{2}{\omega_0^2}(p_{s,0} + \delta p_0)^2 - 1$$
$$\longrightarrow \delta p_0 = \frac{\omega_0^2}{2p_{s,0}}w = \frac{\omega_0}{2\sqrt{2}}w$$
$$\longrightarrow \delta p_1 = \frac{\omega_0}{2\sqrt{2}\cos(\omega_1 T_{s,1})}w.$$

Let T_0 be the period of motion along the trajectory considered (the time interval between sequential passages of the stable phase $x_0 = 0.5$).

Assuming $x = 1$, we find from equation (A1.5)

$$\tanh\left[\omega_2\left(\frac{T_0}{2} - T_{s,1}\right)\right] = \frac{1-d}{2}\frac{\omega_2}{p_1} \approx 1 - \frac{\delta p_1}{p_{s,1}} = \xi.$$

When deriving this equality, we took into account the relationship $p_{s,1}/\omega_0 = \sqrt{(1-d)/2}$, which follows from the second row of (A1.1) for $x_{+1} = (1 + d)/2$ (see also (A1.4)). The period of motion near the separatrix is given by

$$T_0(\omega) = 2T_{s,1} + \frac{1}{\omega_2}\ln\frac{1+\xi}{1-\xi}$$
$$= 2T_{s,1} + \frac{1}{\omega_2}\ln\frac{4\sqrt{1-d}\cos(\omega_1 T_{s,1})}{|w|}. \quad (A1.9)$$

With $d = 1/2$, this formula exactly coincides with formula (9) in [11].

APPENDIX 2

ESTIMATES OF THE MELNIKOV–ARNOLD INTEGRAL

Let us write Melnikov–Arnold integral (13) for the upper perturbation harmonic ($\Omega > 0$) in expanded form:

$$W_{MA} = -\int_{-\infty}^{\infty} p_s[\cos(2\pi m x_s)\cos(\Omega t) \quad (A2.1)$$
$$+ \sin(2\pi m x_s)\sin(\Omega t)]dt.$$

Following [12], the integral will be taken by parts. At each step, we integrate terms with Ω and differenti-

ate the remaining terms. At the third step, terms with Ω^{-3} and the third-order time derivative of p_s will appear under the integral sign. The third-order derivative generates δ_1 functions at the singular points $x_{-1} = (1 - d)/2$ and $x_{+1} = (1 + d)/2$:

$$\frac{d^3 p_s}{dt^3} = \frac{d^3 p_s}{dx^3}p_s^3$$
$$= p_s^3\frac{2}{d(1-d)}[\delta_1(x - x_{+1}) - \delta_1(x - x_{-1})].$$

In this expression, only the principal term with the δ_1 functions is left and formula (4) for the step of the derivatives is used. At the singular points, $p_s/\omega_0 = \sqrt{(1-d)/2}$ (the second row in (A1.1); accordingly, for the first term in (A2.1), we have

$$-\frac{\lambda^{-3}}{8d}\sqrt{\frac{1-d}{2}}[\cos\pi m(1+d)\sin(\lambda\psi_{s,1})$$
$$- \cos\pi m(1-d)\sin(-\lambda\psi_{s,1})]$$
$$= (-1)^{m-1}\frac{\lambda^{-3}}{4d}\sqrt{\frac{1-d}{2}}\cos(\pi md)\sin(\lambda\psi_{s,1}),$$

where $\lambda = \Omega/2\omega_0$ and $\psi_{s,1} = 2\omega_0 T_{s,1}$.

Applying a similar procedure to the second term of (A2.1) and adding up the final results, we obtain an estimate of the Melnikov–Arnold integral for the upper harmonic:

$$W_{MA}(\lambda > 0) \approx (-1)^m\frac{\lambda^{-3}}{4d}\sqrt{\frac{1-d}{2}}\sin(\pi md - \lambda\psi_{s,1}). \quad (A2.2)$$

Accordingly, for the lower perturbation harmonic ($\lambda < 0$), we find

$$W_{MA}(\lambda < 0) \approx (-1)^m\frac{|\lambda|^{-3}}{4d}$$
$$\times \sqrt{\frac{1-d}{2}}\sin(\pi md + |\lambda|\psi_{s,1}). \quad (A2.3)$$

This harmonic makes the same (up to a phase shift) contribution to the separatrix map amplitude as the upper harmonic, which comes into conflict with the numerical experiment. It turns out that the difference between the contributions from the upper and lower harmonics appears only in the next ($\sim\lambda^{-4}$) order of smallness. To find this difference, one must integrate by parts four (not three) times. It is easy to check that, in this case, two terms containing the expressions $p_s(d^4 x_s/dt^4)$ and $\dot{x}_s(d^3 p_s/dt^3)$, which make equal contributions to W_{MA}, must be taken into consideration.

Eventually, for the upper and lower harmonics, we find

$$W_{\text{MA}}(\lambda > 0) \approx (-1)^m \frac{\lambda^{-3}}{4d} \sqrt{\frac{1-d}{2}}$$

$$\times \left[1 + \pi \lambda^{-1} \sqrt{\frac{1-d}{2}} \right] \sin(\pi m d - \lambda \psi_{s,1}) \tag{A2.4}$$

and

$$W_{\text{MA}}(\lambda < 0) \approx (-1)^m \frac{|\lambda|^{-3}}{4d} \sqrt{\frac{1-d}{2}}$$

$$\times \left[1 - \pi |\lambda|^{-1} \sqrt{\frac{1-d}{2}} \right] \sin(\pi m d + |\lambda| \psi_{s,1}), \tag{A2.5}$$

respectively.

These expressions are discussed in Section 2.

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
2. A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics* (Springer, New York, 1992).
3. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics* (Nauka, Moscow, 1988).
4. B. V. Chirikov, Chaos, Solitons, and Fractals **1**, 79 (1991).
5. V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **122**, 175 (2002) [JETP **95**, 154 (2002)].
6. M. Henon and J. Wisdom, Physica D **8**, 157 (1983).
7. S. Bullett, Commun. Math. Phys. **107**, 241 (1986).
8. M. Wojtkowski, Commun. Math. Phys. **80**, 453 (1981); Ergod. Theory & Dyn. Syst. **2**, 525 (1982).
9. V. V. Vecheslavov, Dynamics of Sawtooth Maps: 1. New Numerical Results, *Preprint of the Budker Institute of Nuclear Physics (Siberian Division, Russian Academy of Sciences)*, Novosibirsk, 2000, no. 2000-27; nlin.CD/0005048.
10. B. V. Chirikov and V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **122**, 647 (2002) [JETP **95**, 560 (2002)].
11. L. V. Ovsyannikov, Private Communication, May, 1999.
12. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **119**, 853 (2001) [JETP **92**, 744 (2001)].
13. V. V. Vecheslavov, *Preprint of the Budker Institute of Nuclear Physics* (Siberian Division, Russian Academy of Sciences), Novosibirsk, 1999, no. 99-69.
14. V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **120**, 740 (2001) [JETP **93**, 649 (2001)].
15. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].
16. V. V. Vecheslavov, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 989 (1996) [JETP Lett. **63**, 1047 (1996)].
17. V. V. Vecheslavov, Physica D **131**, 55 (1999).
18. G. M. Zaslavskiĭ and N. N. Filonenko, Zh. Éksp. Teor. Fiz. **54**, 1590 (1968) [Sov. Phys. JETP **27**, 851 (1968)].

Translated by V. Isaakyan

THEORETICAL AND MATHEMATICAL PHYSICS

Asymptotics of Collision Integral Matrix Elements in the Isotropic Case

É. A. Tropp*, L. A. Bakaleĭnikov*, A. Ya. Énder*, and I. A. Énder**

* Ioffe Physicotechnical Institute, Russian Academy of Sciences,
Politekhnicheskaya ul. 26, St. Petersburg, 194021
e-mail: bakal@ammp.ioffe.rssi.ru
** St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 198904 Russia

Received February 20, 2003

Abstract—The method of nonlinear moments, when used to solve the Boltzmann equation, necessitates the calculation of collision integral matrix elements. The matrix elements are hard to calculate numerically, especially at large indices. The asymptotics of the matrix elements are constructed. In terms of the model of pseudopower particle interaction, a formula free of summation is derived. This makes it possible to find the asymptotic behavior of linear and nonlinear elements when two indices are large. For an arbitrary interaction cross section, asymptotic expansions of linear and nonlinear matrix elements in one index are obtained. For Maxwellian molecules, asymptotic formulas are derived for three large indices. © *2003 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

In solving the Boltzmann equation by the method of nonlinear moments, the distribution function is represented as the expansion in basis functions and then a set of differential equations for expansion coefficients is solved. However, with such an approach, one faces difficulties when calculating the matrix of the moments of the nonlinear collision integral. With large indices, associated formulas become so awkward that the matrix elements (MEs) are impossible to calculate with modern computing facilities. For example, formulas containing sixfold summation have been derived for the isotropic (in terms of velocities) Boltzmann equation even under the assumption that the scattering cross section is independent of angles of scattering [1]. In [2], formulas with quadruple summation were derived for an arbitrary angular dependence of the cross section. In calculating with the expressions obtained [1, 2], the machine time rapidly grows and the accuracy drops with increasing ME index.

In [3], recurrence relations for the matrix elements were obtained for the isotropic Boltzmann equation under the assumption that the collision integral is invariant under basis. It turned out that nonlinear MEs can be expressed through linear ones. It was shown [4] that the same is true for the Boltzmann equation that is axisymmetric in terms of velocities, when the distribution function depends on two variable velocities. The authors of [4] also studied the structure of the collision operator and ME properties for the spherically symmetric interaction potential. They determined that unique set of indices that provide nonzero nonlinear MEs. Furthermore, it was found that the vanishing of a number

of nonlinear MEs results in additional relations between isotropic and nonisotropic linear elements, so that the latter can be expressed through the former. This means that all nonlinear axisymmetric MEs can be found through linear isotropic ones. That is why the study of MEs for the isotropic case is of particular importance.

The approach to calculating MEs using recurrence relations [3, 4] has made it possible to significantly cut the computing time and, accordingly, to compute MEs with indices up to 128. MEs with still greater indices can be computed with an asymptotic approach. It also allows one to trace the ME behavior when indices tend to infinity and study the dependence of MEs on the scattering cross section of colliding particles.

In this work, we constructed the asymptotics for linear and nonlinear MEs in the isotropic case and derived a formula for linear MEs that is free of summation. This formula follows from a model of pseudopower particle interaction, which assumes that the interaction cross section depends on the velocity in the same way as in the case of power potentials but the angular distribution is isotropic. With this formula, we managed to find the asymptotics of linear and nonlinear MEs when two indices are large. For Maxwellian molecules, the three-index asymptotics of MEs is found.

BASIC RELATIONSHIPS

When the method of moments is applied to solve the Boltzmann equation, the distribution function is expanded in basis functions and the equation is split into a set of equations for the expansion coefficients. If the distribution function is isotropic, Sonine polyno-

mials

$$S_{1/2}^r(x) = \sum_{p=0}^r \frac{\Gamma(r+3/2)x^{2p}}{p!(r-p)!\Gamma(p+3/2)} \qquad (1)$$

are usually taken as basis functions.

These polynomials are orthogonal to each other with the Maxwellian weighting function $M(v, T_0)$:

$$\int M(v, T_0)S_{1/2}^r(mv^2/2kT_0)S_{1/2}^t(mv^2/2kT_0)v^2 dv = v_r\delta_{r,t},$$

$$M(v, T_0) = (m/2kT_0\pi)^{3/2}\exp(-mv^2/2kT_0), \qquad (2)$$

$$v_r = (2r+1)!!/(2r)!!.$$

Here, m is the mass of a particle, T_0 is the gas temperature, k is the Boltzmann constant, and v_r is the squared norm of a Sonine polynomial. In the isotropic case, the expansion of the distribution function has the form

$$f(v, t) = n_0 M(v, T_0)\sum_{r=0}^\infty C_r(t)S_{1/2}^r(mv^2/2kT_0), \qquad (3)$$

where n_0 is the numerical density of the particles.

Substituting this expansion into the Boltzmann equation with the subsequent multiplication of both sides of the resulting expression by $S_{1/2}^r(mv^2/2kT_0)$ and integration over velocities yields

$$dC_r/dt = \sum_{r_1, r_2} K_{r_1, r_2}^r C_{r_1} C_{r_2}. \qquad (4)$$

The matrix elements K_{r_1, r_2}^r are determined through the collision integral $I(f, f)$:

$$K_{r_1, r_2}^r = 4\pi\left(\int_0^\infty S_{1/2}^r I(MS_{1/2}^{r_1}, MS_{1/2}^{r_2})v^2 dv\right)/v_r. \qquad (5)$$

The collision integral $I(f, f)$ is a nonlinear integral operator given by

$$I(f, f) = \int(f(\mathbf{v}_1)f(\mathbf{v}_2) - f(\mathbf{v})f(\mathbf{v}'))g\sigma(g, \Theta)d\mathbf{v}' \cdot d\mathbf{k}. \qquad (6)$$

Here, v, v', v_1, and v_2 are the velocities of colliding particles before and after collision, $g = |\mathbf{g}| = |\mathbf{v} - \mathbf{v}'|$ is the relative velocity of the particles, $\mathbf{k}$ is the unit vector directed along the relative velocity vector after collision, and $\sigma(g, \Theta)$ is the differential scattering cross section. The angle of scattering Θ is found from the relationship $\cos\Theta = \mathbf{k} \cdot \mathbf{g}/|\mathbf{g}|$.

If one of the indices (r_1 or r_2) of the matrix element K_{r_1, r_2}^r goes to zero, the collision integral turns to the linear operator.

Expressions for linear and nonlinear MEs in the case of arbitrary power interaction potentials are given in [3]. The linear MEs $K_{0, n}^r$ and $K_{n, 0}^r$ involve one-tuple

sums, the number of terms in which depend on the minimal value of the indices n and r. The expressions for the nonlinear elements in the case of power interaction potentials involve quadruple sums, which are extremely difficult to calculate when the indices are large. Because of this, it was suggested [3] that the elements K_{r_1, r_2}^r be calculated with the basic relationship between the matrix elements, which follows from the invariance of the collision integral under basis. This relationship has the form

$$K_{p, n}^r = \frac{1}{p}[(n + p - r - 1 - \mu)L_{p-1, n}^r + rK_{p-1, n}^{r-1} - (n+1)K_{p-1, n+1}^r] \qquad (7)$$

and allows one to sequentially find MEs with an increasing first subscript. A similar relationship can be written for MEs with an increasing second subscript:

$$K_{n, p}^r = \frac{1}{p}[(n + p - r - 1 - \mu)L_{n, p-1}^r + rK_{n, p-1}^{r-1} - (n+1)K_{n+1, p-1}^r]. \qquad (8)$$

Clearly, from relationships (7) and (8), one can find nonlinear MEs with arbitrary indices provided that the linear elements $K_{0, n}^r$ and $K_{n, 0}^r$ are known.

This computing scheme was applied in [3] to calculate nonlinear MEs and allowed us to raise the maximal number N_0 of expansion terms up to $N_0 = 128$. It turned out that the calculation of large-index nonlinear MEs requires linear elements to be known with a high accuracy. To meet this requirement, precision arithmetic procedures that ensure calculations accurate to a high decimal place were applied [3]. The final error in MEs with $N_0 = 128$ were no larger than 10^{-8}. This makes it possible to directly check the convergence of asymptotic formulas. The construction of the ME asymptotics allows one to extend the computing procedure to the range of very large indices.

ASYMPTOTICS FOR LINEAR MATRIX ELEMENTS

The expressions for linear MEs have the form [3]

$$K_{0, n}^r = \left(\frac{4kT_0}{m}\right)^\mu \frac{r!}{\Gamma(r+3/2)2^{r+n}}$$

$$\times \sum_{q=1}^{\min(n, r)} 2^{2q}W_{r, q}^n\Gamma(q + \mu + 3/2)J_q/q!, \qquad (9)$$

$$K_{n, 0}^r = \left(\frac{4kT_0}{m}\right)^\mu \frac{r!}{\Gamma(r+3/2)2^{r+n}}$$

$$\times \sum_{q=1}^{\min(n, r)} 2^{2q}W_{r, q}^n\Gamma(q + \mu + 3/2)\tilde{J}_q/q!. \qquad (10)$$

Here,

$$W^n_{r,q} = \frac{-\mu(-\mu+1)\ldots(-\mu+n+r-2q-1)}{(n-q)!(r-q)!}$$

$$= \frac{\Gamma(-\mu+n+r-2q)}{\Gamma(-\mu)(n-q)!(r-q)!}. \tag{11}$$

The above expressions were obtained for the case of power interaction potentials when the angular and velocity dependences of the cross section can be separated; that is,

$$\sigma(g, \Theta) = \frac{1}{4\pi} g^{\gamma-1} F_\gamma((\sin\Theta/2)^2). \tag{12}$$

The parameter μ in (9) and (10) and γ are related as $\mu = \gamma/2$, and the coefficients J_q and $\tilde{J}_q$ are the integrals of the angular component $F_\gamma(z)$ with the weights z^q and $((1-z)^q - 1)$, respectively:

$$J_q = 4\pi \int_0^1 F_\gamma(z) z^q dz,$$

$$\tilde{J}_q = 4\pi \int_0^1 F_\gamma(z)((1-z)^q - 1) dz. \tag{13}$$

Works concerned with the kinetic theory of gases often use the model of pseudopower potentials of particle interaction, to which differential cross section (10) with $F(z) = 1/4\pi$ corresponds; in other words, particle

scattering is assumed to be isotropic. Then,

$$J_q = \frac{1}{q+1}, \quad \tilde{J}_q = \frac{1}{q+1} - 1. \tag{13a}$$

In this case, expression (9) for linear MEs can be transformed into a form that is much more convenient for asymptotics construction and numerical calculation. Before making this transformation, we note that the sum in (9) in view of (13a) takes the form

$$S = \sum_{q=1}^{\min(n,r)} 2^{2q} W^n_{r,q} \Gamma(q+\mu+3/2) J_q/q!$$

$$= \sum_{q=1}^{\min(n,r)} 2^{2q} \frac{\Gamma(q+\mu+3/2)\Gamma(-\mu+n+r-2q)}{(n-q)!(r-q)!\Gamma(-\mu)(q+1)!}. \tag{14}$$

The factorials entering into the denominator of the sum are rearranged as

$$(r-q)! = \frac{r!}{(r-q+1)\ldots(r-1)r}$$

$$= \frac{r!}{(-1)^q(-r)(-r+1)\ldots(-r+q-1)} = \frac{(-1)^q r!}{(-r)_q}, \tag{15}$$

$$(n-q)! = \frac{(-1)^q n!}{(-n)_q}. \tag{16}$$

Here, $(a)_q$ are Pochhammer symbols

$$(a)_0 = 1, \quad (a)_q = a(a+1)\ldots(a+q-1);$$

$$q = 1, 2, 3, \ldots, \tag{17}$$

and the function Γ in the denominator of sum (14) can be expressed as

$$\Gamma(-\mu+n+r-2q) = \frac{\pi}{\sin(\pi\mu)(-1)^{n+r+1}\Gamma(1+\mu-n-r+2q)}$$

$$= \frac{\pi^{3/2}}{\sin(\pi\mu)(-1)^{n+r+1}2^{\mu-n-r+2q}\Gamma(q+(\mu+1-n-r)/2)\Gamma(q+(\mu+1-n-r)/2+1/2)}. \tag{18}$$

Substituting (15), (16), and (18) into (14) and taking into account that $(a)_p = \Gamma(a+p)/\Gamma(a)$, we find

$$S = \frac{\pi\Gamma(\mu+1/2)(-1)^{r+n+1}}{4G(-\mu)(n+1)!(r+1)!\sin(\pi\mu)\Gamma(\mu-1-r-n)}$$

$$\times \sum_{p=2}^{\min(n,r)+1} \frac{(\mu+1/2)_p(-n-1)_p(-r-1)_p}{((\mu-1-r-n)/2)_p((\mu-1-r-n)/2+1/2)_p p!}. \tag{19}$$

Now we note that

$$\sum_{p=0}^{\infty} \frac{(a)_p(b)_p(c)_p}{(d)_p(e)_p p!} z^p = {}_3F_2\left(\begin{matrix} a, b, c \\ d, e \end{matrix}\bigg| z\right). \tag{20}$$

If one of the superscripts is a negative integer, the series contains a finite number of terms and the sum in (19) can be represented as

$$s = \sum_{p=2}^{\min(n,r)+1} \frac{(\mu+1/2)_p(-n-1)_p(-r-1)_p}{((\mu-1-r-n)/2)_p((\mu-1-r-n)/2+1/2)_p p!}$$

$$= {}_3F_2\left(\begin{matrix} \mu+1/2-n-1-r-1 \\ (\mu-1-n-r)/2(\mu-1-n-r)/2+1/2 \end{matrix}\bigg| z=1\right) \tag{21}$$

$$- 1 - \frac{4(\mu+1/2)(-n-1)(-r-1)}{(\mu-1-r-n)(\mu-r-n)}.$$

According to the Saalschutz theorem [5], if $c = -m$ (m is an integer) and $e + d = a + b + c + 1$, the hypergeometric function ${}_3F_2\left(\begin{matrix} a, b, c \\ d, e \end{matrix}\bigg| z = 1\right)$ is given by

$${}_3F_2\left(\begin{matrix} a, b, -m \\ d, 1+a+b-d-m \end{matrix}\bigg| z = 1\right) = \frac{(d-a)_m(d-b)_m}{(d)_m(d-a-b)_m}. \tag{22}$$

In the case considered, this relationship is met and

$$s = \frac{(-\mu/2 - 1 - n/2 - r/2)_{R+1}(\mu/2 - 1/2 - n/2 - r/2 + R_1 + 1)_{R+1}}{(\mu/2 - 1/2 - n/2 - r/2)_{R+1}(-\mu/2 - n/2 - r/2 + R_1)_{R+1}}$$
$$- 1 - \frac{4(\mu + 1/2)(-n - 1)(-r - 1)}{(\mu - n - r)(\mu - n - r - 1)}. \tag{23}$$

Here, $R = \min(n, r)$ and $R_1 = \max(n, r)$. Taking into account (23) and (19), we find that the linear ME $K_{0,n}^r$ is given by

$$K_{0,n}^r = \left(\frac{4kT_0}{m}\right)^\mu \frac{r! n_0^2}{\Gamma(r + 3/2)2^{r+n+2}}$$
$$\times \frac{\Gamma(\mu + 1/2)\Gamma(r + n + 2 - \mu)}{\Gamma(-\mu)(n + 1)!(r + 1)!}$$
$$\times \left[\frac{((-\mu - n - r)/2 - 1)_{R+1}((\mu - n - r)/2 + R_1 + 1/2)_{R+1}}{((\mu - 1 - n - r)/2)_{R+1}((-\mu - n - r)/2 + R_1)_{R+1}}\right.$$
$$\left. - 1 - \frac{4(\mu + 1/2)(n + 1)(r + 1)}{(\mu - n - r)(\mu - n - r - 1)}\right]. \tag{24}$$

Expression (24) is the desired representation of the linear ME.

Numerical calculations show that linear MEs with slightly differing indices r and n are the greatest; consequently, it is precisely in this case that the asymptotics is of greatest interest. Expression (24) allows one to construct such an asymptotics. Let us assume that $r \longrightarrow \infty$ and $n \longrightarrow \infty$, so that the difference between the indices remains constant; that is, $R \longrightarrow \infty$, $R_1 \longrightarrow \infty$, and $\Delta = R_1 - R$ is fixed. We will first find the asymptotics of the fraction Z_1, which contains Pochhammer symbols in the brackets of (24). Taking into account that $r + n = 2R + \Delta$ and using the relationship

$$(\alpha - (R + 1))_{R+1} = (\alpha - (R + 1))(\alpha - (R + 1) + 1)$$
$$\ldots(\alpha - (R + 1) + R) = (-1)^{R+1}(-\alpha + 1)(-\alpha + 2) \tag{25}$$
$$\ldots(-\alpha + (R + 1)) = (-1)^{R+1}(-\alpha + 1)_{R+1},$$

we transform Z_1 into the form

$$Z_1 = \frac{((-\mu - n - r)/2 - 1)_{R+1}((\mu - n - r)/2 + R_1 + 1/2)_{R+1}}{((\mu - 1 - n - r)/2)_{R+1}((-\mu - n - r)/2 + R_1)_{R+1}}$$
$$= \frac{((\mu + \Delta)/2 + 1)_{R+1}((\mu + \Delta + 1)/2)_{R+1}}{((-\mu + \Delta)/2)_{R+1}((-\mu + \Delta + 1)/2)_{R+1}}. \tag{26}$$

In going from Pochhammer symbols to the ratio of the Γ functions and using the asymptotics of the ratio of large-index Γ functions (see, e.g., [5])

$$\frac{\Gamma(z + \alpha)}{\Gamma(z + \beta)} = z^{\alpha - \beta}\left(1 + \frac{1}{2z}(\alpha - \beta)(\alpha + \beta - 1) + O(z^{-2})\right), \tag{27}$$
$$z \longrightarrow \infty,$$

we come to

$$Z_1 = 2^{2\mu + 1}\frac{\Gamma(-\mu + \Delta)}{\Gamma(\mu + \Delta + 1)}(R + 1)^{2\mu + 1}$$
$$\times \left[1 + \frac{1}{2(R + 1)}(2\mu + 1)\Delta + O((R + 1)^{-2})\right], \tag{28}$$
$$R \longrightarrow \infty.$$

The asymptotics of the second fraction, Z_2, in the brackets of (24) is obvious:

$$Z_2 = \frac{4(\mu + 1/2)(n + 1)(r + 1)}{(\mu - n - r)(\mu - n - r - 1)}$$
$$= \frac{(\mu + 1/2)(R + 1)(R + \Delta + 1)}{((\mu - \Delta)/2 - R)((\mu - \Delta)/2 - R - 1/2)} \tag{29}$$
$$= (\mu + 1/2)\left[1 + \frac{\mu + 3/2}{R + 1} + O((R + 1)^{-2})\right], \quad R \longrightarrow \infty.$$

The asymptotics of the coefficient multiplying the brackets in (24) is easily found by applying the duplication formula to the Γ function in the numerator and formula (27) for the ratio of Γ functions:

$$U = \frac{r! G(r + n + 2 - \mu)}{\Gamma(r + 3/2)2^{r+n+2}(n + 1)!(r + 1)!} = 2^{-\mu - 1}$$
$$\times (R + 1)^{-\mu - 2}\frac{1}{\sqrt{\pi}}\left[1 + \frac{B}{2(R + 1)} + O((R + 1)^{-2})\right],$$
$$R \longrightarrow \infty,$$

$$B = ((-\mu + \Delta) - 2)(-\mu + \Delta)/4$$
$$+ (-\mu - \Delta - 1)(-\mu + 3\Delta + 1)/4 - (\Delta_1 - 1/4),$$

$$\Delta_1 = \begin{cases} 0 & r \le n \\ \Delta & r > n. \end{cases} \tag{30}$$

Taken together, (28)–(30) yield a final expression for the linear element asymptotics:

$$K_{0,n}^r = P(\mu)\left\{ 2^\mu \frac{\Gamma(-\mu + \Delta)}{\Gamma(\mu + \Delta + 1)}(R+1)^{\mu-1} \right.$$

$$\times \left[1 + \frac{A}{2(R+1)} + O((R+1)^{-2}) \right] - 2^{-\mu-1}(\mu + 3/2)$$

$$\left. \times (R+1)^{-\mu-2}\left[1 + \frac{B + 2\mu + 1}{2(R+1)} + O((R+1)^{-2}) \right] \right\}, \tag{31}$$

$$R \longrightarrow \infty,$$

$$P(\mu) = \left(\frac{4kT_0}{m} \right)^\mu \frac{\Gamma(\mu + 1/2)n_0^2}{\Gamma(-\mu)\sqrt{\pi}},$$
$$A = B + \Delta(2\mu + 1).$$

From (31), it follows that the asymptotics of $K_{0,n}^r$ has two parts with the principal terms behaving as $(R+1)^{\mu-1}$ and $(R+1)^{-\mu-2}$, respectively. Which of them specifies the asymptotic behavior depends on μ: with $\mu > -1/2$, the first term dominates; otherwise, the second prevails. With $\mu = 0$, the found terms of the asymptotics vanish if $\Delta \ne 0$. If $\Delta = 0$, we have

$$K_{0,r}^r = n_0^2(r+1)^{-1}(1 + O((r+1)^{-2})). \tag{32}$$

This result corresponds to the well-known formula for a linear element in the case of pseudo-Maxwellian molecules:

$$K_{0,n}^r = \frac{n_0^2}{r+1}\delta_{r,n}. \tag{33}$$

Now we will find the asymptotics of linear MEs in indices n and r for the general case, i.e., for nonisotropic differential cross section (12) of particle interaction. Consider first the asymptotics of $K_{0,n}^r$ at $r \longrightarrow \infty$ and fixed n. Using formula (27) for the ratio of the large-index Γ function, we find from (11) that

$$W_{r,q}^n = \frac{1}{\Gamma(-\mu)(n-q)!}r^{n-q-\mu-1}$$

$$\times \left[1 + \frac{1}{2r}(n-q-\mu-1)(n-3q-\mu) + O(r^{-2}) \right], \tag{34}$$

$$r \longrightarrow \infty.$$

From (32), it follows that the sum in the expression for $K_{0,n}^r$ is an asymptotic expansion and that the major contribution is due to the term with $q = 1$. Substituting (34) into (9) in view of (27), we obtain for the first terms in the expansion of $K_{0,n}^r$

$$K_{0,n}^r = \left(\frac{4kT_0}{m} \right)^\mu \frac{n_0^2}{2^{n-2}} \frac{\Gamma(\mu + 5/2)}{\Gamma(-\mu)(n-1)!}2^{-r}r^{n-\mu-5/2}$$

$$\times \left\{ J_1 + \frac{1}{r}\left[J_1\left(\frac{(n-\mu-2)(n-\mu-3)}{2} - \frac{3}{8} \right) \right. \right. \tag{35}$$

$$\left. \left. + 2J_2(m + 5/2)(n-1) \right] + O(r^{-2}) \right\}, \quad r \longrightarrow \infty.$$

The asymptotics in terms of the second index can be found in a similar way:

$$W_{r,q}^n = \frac{\Gamma(-\mu + n + r - 2q)}{\Gamma(-\mu)(r-q)!\Gamma(n-q+1)} = \frac{1}{\Gamma(-\mu)(r-q)!}$$

$$\times n^{-\mu+r-q-1}$$

$$\times \left[1 + \frac{1}{2n}(r-\mu-q-1)(r-\mu-3q) + O(n^{-2}) \right] \tag{36}$$

$$n \longrightarrow \infty,$$

which gives

$$K_{0,n}^r = \left(\frac{4kT_0}{m} \right)^\mu \frac{r!n_0^2}{\Gamma(r+3/2)2^{r-2}} \frac{\Gamma(\mu + 5/2)}{\Gamma(-\mu)(r-1)!}$$

$$\times 2^{-n}n^{r-\mu-2}\left\{ J_1 + \frac{1}{n}\left[J_1 \frac{(r-\mu-2)(r-\mu-3)}{2} \right. \right. \tag{37}$$

$$\left. \left. + 2J_2(\mu + 5/2)(r-1) \right] + O(n^{-2}) \right\}, \quad n \longrightarrow \infty.$$

The asymptotics of linear MEs with another sequence of subscripts can be found from (35) and (37) with $\tilde{J}_q$ substituted for J_q. This follows from comparison (9) and (10).

ASYMPTOTICS OF NONLINEAR MATRIX ELEMENTS

The results obtained in the previous section, together with recurrence relations (7) and (8), allow us to construct the asymptotics of the nonlinear elements $K_{p,n}^r$. In terms of the pseudopower interaction model, one can find the behavior of $K_{p,n}^r$ with fixed p and large r and n, substituting asymptotics (31) into recurrence relation (7). It should be borne in mind here that $R = \min(r, n)$ and that it is necessary to consider two cases

in order to calculate the result of substitution. In the first case, $r > n$, which yields $R = n$ and $\Delta = |r - n| = r - n \geq 1$. In the second case, $r \leq n$ and, accordingly, $R = r$ and $\Delta = n - r \geq 0$. Passing in the recurrence relation from the variables r and n to R and Δ, we find for these cases

$$
K_{1,n}^r = P(\mu)\left\{ \frac{\Gamma(-\mu + \Delta - 1)}{\Gamma(\mu + \Delta)} 2^\mu (R+1)^{\mu-1} \right.
$$
$$
+ O((R+1)^{\mu-2}) - (\mu + 3/2)2^{-\mu-1}(R+1)^{-\mu-2} \quad (38a)
$$
$$
\left. + O((R+1)^{-\mu-3}) \right\}, \quad r > n,
$$

$$
K_{1,n}^r = P(\mu)\left\{ \frac{\Gamma(-\mu + \Delta + 1)}{\Gamma(\mu + \Delta + 2)} 2^\mu (R+1)^{\mu-1} \right.
$$
$$
+ O((R+1)^{\mu-2}) - (\mu + 3/2)2^{-\mu-1}(R+1)^{-\mu-2} \quad (38b)
$$
$$
\left. + O((R+1)^{-\mu-3}) \right\}, \quad r \leq n.
$$

To find the asymptotics of elements $K_{2,n}^r$, one must consider three cases, namely, $r > n + 1$, $r = n + 1$, and $r < n + 1$. This is because for $K_{1,m}^q$, which enter into recurrence relation (7) with $p = 2$, it is necessary to use (38a) or (38b) depending on the relationship between q and m. This yields

$$
K_{2,n}^r = P(\mu)\left\{ \frac{\Gamma(-\mu + \Delta - 2)}{\Gamma(\mu + \Delta - 1)} 2^\mu (R+1)^{\mu-1} \right.
$$
$$
+ O((R+1)^{\mu-2}) - (\mu + 3/2)2^{-\mu-1}(R+1)^{-\mu-2} \quad (39a)
$$
$$
\left. + O((R+1)^{-\mu-3}) \right\}, \quad r > n + 1,
$$

$$
K_{2,n}^r = P(\mu)\left\{ \frac{\Gamma(-\mu + 1)}{\Gamma(\mu + 2)} 2^\mu (R+1)^{\mu-1} \right.
$$
$$
+ O((R+1)^{\mu-2}) - (\mu + 3/2)2^{-\mu-1}(R+1)^{-\mu-2} \quad (39b)
$$
$$
\left. + O((R+1)^{-\mu-3}) \right\}, \quad r = n + 1,
$$

$$
K_{2,n}^r = P(\mu)\left\{ \frac{\Gamma(-\mu + \Delta + 2)}{\Gamma(\mu + \Delta + 3)} 2^\mu (R+1)^{\mu-1} \right.
$$
$$
+ O((R+1)^{\mu-2}) - (\mu + 3/2)2^{-\mu-1}(R+1)^{-\mu-2} \quad (39c)
$$
$$
\left. + O((R+1)^{-\mu-3}) \right\}, \quad r < n + 1.
$$

Now formulas (31), (38), and (39) for the asymptotics of elements $K_{p,n}^r$ at $p = 0$, 1, and 2 can be represented in the form

$$
K_{p,n}^r = P(\mu)\left\{ \frac{\Gamma(-\mu + |r - n - p|)}{\Gamma(\mu + |r - n - p| + 1)} 2^\mu (R+1)^{\mu-1} \right.
$$
$$
+ O((R+1)^{\mu-2}) - (\mu + 3/2)2^{-\mu-1}(R+1)^{-\mu-2} \quad (40)
$$
$$
\left. + O((R+1)^{-\mu-3}) \right\}.
$$

One can prove by induction that this formula is also valid for $p > 2$. Three cases should be considered in this case: $r \geq n + p + 1$, $n < r < n + p + 1$, and $r \leq n$. Such a selection of the index ranges stems from the need to remove the modulus sign in the arguments of the Γ functions and also takes into account the r–R and n–R relations. Formula (40) represents the ME asymptotics for fixed p and $\Delta = |r - n|$, as well as for large r and n, for the case of pseudopower particle interaction.

The asymptotics of nonlinear MEs in one index with the others fixed can be found by the same expedient. The asymptotics of the ME $K_{0,n}^r$ with n fixed and $r \longrightarrow \infty$ has the form

$$
K_{0,n}^r = \tilde{P}(\mu) \frac{r^{n-\mu-5/2}}{2^{n+r-2}(n-1)!}
$$
$$
\times \left\{ J_1 + \frac{1}{r}C_0(n, \mu) + O(r^{-2}) \right\}, \quad (41)
$$

where

$$
\tilde{P}(\mu) = \left(\frac{4kT_0}{m} \right)^\mu \frac{\Gamma(\mu + 5/2)n_0^2}{\Gamma(-\mu)}, \quad (41a)
$$

$$
C_0(n, \mu) = J_1\left(\frac{(n - \mu - 2)(n - \mu - 3)}{2} - \frac{3}{8} \right)
$$
$$
+ 2J_2(\mu + 5/2)(n - 1). \quad (42)
$$

Substituting (41) into recurrence relation (7) and separating out the principal terms in r, we find the asymptotics of $K_{1,n}^r$. For any finite r, the asymptotics of elements $K_{p,n}^r$ is found by repeating this procedure. One can show by induction that

$$
K_{p,n}^r = \tilde{P}(\mu)\left\{ (n - p) \frac{r^{n-\mu+p-5/2}}{2^{n+r+p-2}n!p!} J_1 \right.
$$
$$
\left. + \frac{r^{n-\mu+p-7/2}}{2^{n+r+p-2}(n-1)!} C_p(n, \mu) + O\left(\frac{r^{n-\mu+p-9/2}}{2^r} \right) \right\}. \quad (43)
$$

The coefficients $C_p(n, \mu)$ are related as

$$C_p(n, \mu) = \frac{2}{p}\left(C_{p-1}(n, \mu) - \frac{n+1}{2n}C_{p-1}(n+1, \mu)\right)$$

$$- \frac{2}{np!}(n - \mu + p - 6)(n - p + 1)J_1. \quad (44)$$

Prior to finding $C_p(n, \mu)$ in explicit form, we note that $C_0(n, \mu)$ as a function of n is a linear combination of power functions n^i ($i = 0, 1, 2$) and the operator

$$L(F(n)) = F(n) - \frac{n+1}{2n}F(n+1)$$

transforms n^i (at $i \geq 0$) into a linear combination like

$$\sum_{j=-1}^{i} T_j^i n^j.$$

Taking into consideration that $L(n^{-1}) = (1/2)n^{-1}$ and that the free term in (44) is a combination of the same power functions of the index n, we arrive at

$$C_p(n, \mu) = \frac{A_2}{p!}F_2(n) + \frac{1}{p!}F_1(n)[A_1 - 3pA_2 - 2pJ_1]$$

$$+ \frac{1}{p!}F_0(n)[A_0 - 2pA_1 + 3p(p-2)A_2 + 2J_1p(p+\mu+4)]$$

$$- \frac{1}{(p-1)!}F_{-1}(n)[A_0 + A_2 - A_1(p-2)$$

$$+ A_2(p-1)(p-5) + 2J_1(p-1)(\mu+4)], \quad (45)$$

where

$$A_2 = J_1/2, \quad A_1 = (\mu + 5/2)(2J_2 - J_1),$$

$$A_0 = \frac{J_1}{2}\left(\mu+2)(\mu+3) - \frac{3}{4}\right) - 2J_2(\mu + 5/2). \quad (46)$$

Note that formula (43) applies only if $p \leq n$. With $p = n$, the order of the principal term changes and the general sequence of constructing $K_{n+1, n}^r$ by using recurrence relation (7) breaks. The asymptotics of elements with another sequence of subscripts is easily found from asymptotics (43) and (45) by substituting $\tilde{J}_q$ for J_q. Indeed, this is true for linear elements $K_{0, n}^r$ and $K_{n, 0}^r$ and the recurrence relations for finding $K_{n, p}^r$ through $K_{n, p-1}^r$, $K_{n, p-1}^{r-1}$, and $K_{n+1, p-1}^r$ coincide with the recurrence relations for finding $K_{p, n}^r$ through $K_{p-1, n}^r$, $K_{p-1, n}^{r-1}$, and $K_{p-1, n+1}^r$. Therefore,

$$K_{p, n}^r = \tilde{P}(\mu)\left\{(n-p)\frac{r^{n-\mu+p-5/2}}{2^{n+r+p-2}n!p!}\tilde{J}_1\right.$$

$$\left. + \frac{r^{n-\mu+p-7/2}}{2^{n+r+p-2}(n-1)!}\tilde{C}_p(n, \mu) + O\left(\frac{r^{n-\mu+p-9/2}}{2^r}\right)\right\}, \quad (47)$$

$$\tilde{C}_p(n, \mu) = \frac{\tilde{A}_2}{p!}F_2(n) + \frac{1}{p!}F_1(n)[\tilde{A}_1 - 3p\tilde{A}_2 - 2p\tilde{J}_1]$$

$$+ \frac{1}{p!}F_0(n)[\tilde{A}_0 - 2p\tilde{A}_1 + 3p(p-2)\tilde{A}_2 + 2\tilde{J}_1p(p+\mu+4)]$$

$$- \frac{1}{(p-1)!}F_{-1}(n)[\tilde{A}_0 + \tilde{A}_2 - \tilde{A}_1(p-2)$$

$$+ \tilde{A}_2(p-1)(p-5) + 2\tilde{J}_1(p-1)(\mu+4)], \quad (48)$$

where

$$\tilde{A}_2 = \tilde{J}_1/2, \quad \tilde{A}_1 = (\mu + 5/2)(2\tilde{J}_2 - \tilde{J}_1),$$

$$\tilde{A}_0 = \frac{\tilde{J}_1}{2}\left((\mu+2)(\mu+3) - \frac{3}{4}\right) - 2\tilde{J}_2(\mu + 5/2). \quad (49)$$

This asymptotics is also valid for $p \leq n$. If $p = n$, the expressions for $C_n(n, \mu)$ and $\tilde{C}_n(n, \mu)$ can be represented as

$$C_n(n, \mu) = \frac{4(\mu + 5/2)}{n!}(J_2 - J_1),$$

$$\tilde{C}_n(n, \mu) = \frac{4(\mu + 5/2)}{n!}(\tilde{J}_2 - \tilde{J}_1). \quad (50)$$

This, in view of the relationships $\tilde{J}_1 = -J_1$ and $\tilde{J}_2 = J_2 - 2J_1$, gives $\tilde{C}_n(n, \mu) = C_n(n, \mu)$. In other words, the asymptotics of $K_{n, n}^r$ may be calculated from either of formulas (43) and (47).

With large n, the asymptotics of nonlinear elements $K_{p, n}^r$ is found in the same way. Writing formula (37) in the form

$$K_{0, n}^r = \tilde{P}(\mu)\frac{r}{\Gamma(r + 3/2)}$$

$$\times \frac{n^{r-\mu-2}}{2^{n+r-2}}\left\{J_1 + \frac{1}{n}D_0(r, \mu) + O(n^{-2})\right\}, \quad (51)$$

where

$$D_0(r, \mu) = J_1\frac{(-\mu+r-2)(-\mu+r-3)}{2}$$

$$\times 2J_2(\mu + 5/2)(r-1), \quad (52)$$

and sequentially applying recurrence relation (7), we arrive at

$$K_{p, n}^r = \tilde{P}(\mu)\frac{r}{\Gamma(r+3/2)}\left\{\frac{n^{r+p-\mu-2}}{2^{n+r+p-2}}\frac{1}{p!}J_1\right.$$

$$\left. + \frac{n^{r+p-\mu-3}}{2^{n+r+p-2}}\frac{1}{p!}D_p(r, \mu) + O(n^{r-\mu+p-4})\right\}. \quad (53)$$

$$D_p(r, \mu) = D_{p-1}(r, \mu) + (-\mu - 3r + p)J_1. \quad (54)$$

In view of the fact that $D_1(r, \mu) = (-\mu - 3r)J_1 + 1J_1 + D_0(r, \mu)$, one easily finds the coefficient $D_p(r, \mu)$ in explicit form:

$$D_p(r, \mu) = p(-\mu - 3r)J_1 + \sum_{i=1}^{p} iJ_1 + D_0(r, \mu)$$

$$= p(-\mu - 3r)J_1 + \frac{(p+1)p}{2}J_1 + D_0(r, \mu). \quad (55)$$

It should be noted that the one-index asymptotics found above are valid for any angular dependence of the scattering differential cross section.

ASYMPTOTIC VERSUS EXACT VALUES OF MATRIX ELEMENTS

The ME asymptotics found in the previous sections were compared with the results of numerical calculation. First, we contrasted the calculated results for $K_{0,n}^r$ with analytical expression (24) for a linear ME in the case of pseudopower particle interaction. The values of $K_{0,n}^r$ found numerically and those calculated by formula (24) coincide with a high accuracy for $\mu \in [-1.5, 1]$ and for the indices in the ranges $0 \le r \le 100$ and $0 \le n \le 100$. This supports the validity of the algorithm for $K_{0,n}^r$ computation.

The quality of the asymptotic approximations for pseudopower potentials were estimated with the ratios $\eta(r; n; \mu) = \tilde{K}_{0,n}^r / K_{0,n}^r$, where $\tilde{K}_{0,n}^r$ denotes the sum of explicitly separated terms in asymptotics (35) or (37). The value of $K_{0,n}^r$ was calculated by formula (24). It turned out that an increase in μ slows down, albeit insignificantly, the trend of $K_{0,n}^r$ toward the asymptotic value. For example, the ratio $\eta(5; n; -1.5)$ reaches 0.9 at $n = 18$, while $\eta(5; n; 1)$ reaches 0.9 at $n = 43$. At the same time, the index r influences appreciably the closeness of the asymptotic and exact values of $K_{0,n}^r$. With $r = 3$, the ratio $\eta(r; n; \mu)$ reaches 0.9 at $n = 13$; with $r = 10$, at $n = 100$. Such high sensitivity of $\eta(r; n; \mu)$ to r readily follows from nonuniform asymptotic expansion (37). The same situation takes place for the asymptotics of $K_{0,n}^r$ in index r (see formula (35)). In this case, the rate of approach to the asymptotic value depends strongly on the parameter n.

Of great interest is to trace the difference between the exact expression for the ME $K_{0,n}^r$ and its asymptotic value as the indices r and n grow simultaneously. This case is important because linear MEs take the greatest values near the principal diagonal. Figure 1 shows the r
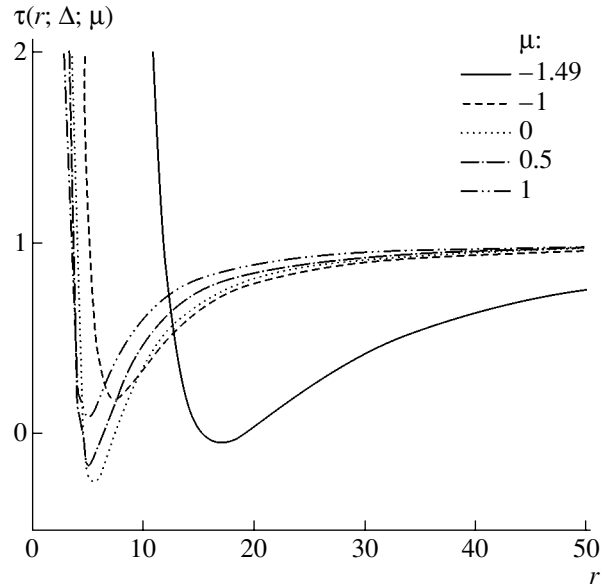
Fig. 1. $\tau(r; \Delta; \mu) = \tilde{K}_{0,\Delta+r}^r / K_{0,\Delta+r}^r$ vs. r for various μ and $\Delta = 5$.

dependences of the ratio $\tau(r; \Delta; \mu) = \tilde{K}_{0,\Delta+r}^r / K_{0,\Delta+r}^r$ for various μ and $\Delta = 5$. Here, as before, $\tilde{K}_{0,\Delta+r}^r$ denotes the sum of explicitly separated terms in asymptotic formula (31). If μ is far from -1.5, its effect on $\tau(r; \Delta; \mu)$ is insignificant. The rate of approach to the asymptotic value increases with increasing μ but slows down as μ approaches -1.5. With $\mu = -1.4999$ and $\Delta = 5$, $\tau(r; \Delta; \mu)$ reaches 0.9 at $r \approx 800$. The Δ dependence of $\tau(r; \Delta; \mu)$ is demonstrated in Fig. 2. It is seen that the value of Δ has a noticeable effect on the ratio $\tilde{K}_{0,\Delta+r}^r / K_{0,\Delta+r}^r$. As Δ increases, the asymptotic and exact values of linear MEs approach each other at larger r.

Asymptotic formulas (35) and (37) were compared with linear MEs and potentials that include the angular dependence of the cross section. In this case, formula (24) is inapplicable and MEs are found by using a computational algorithm. We calculated $K_{0,n}^r$ for a hybrid model where the velocity dependence of the cross section was taken to be the same as in the hard sphere model ($\mu = 0.5$) and the angular part is as that for true Maxwellian molecules [6]:

$$F_M(\Theta)$$

$$= \frac{(\cos 2\varphi)^{1/2}}{4\sin\Theta\sin 2\varphi(\cos^2\varphi \mathbf{K}(\sin\varphi) - \cos 2\varphi \mathbf{E}(\sin\varphi))}. \quad (56)$$

Here, Θ and φ are related as

$$\frac{\pi - \Theta}{2} = (\cos 2\varphi)^{1/2}\mathbf{K}(\sin\varphi),$$

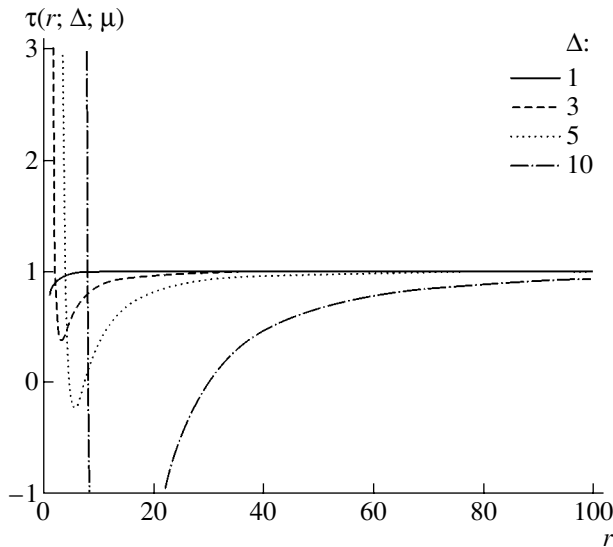and $\mathbf{K}(k)$ and $\mathbf{E}(k)$ are complete elliptic integrals of the first and second kind, respectively.

$\tau(r; \Delta; \mu)$

Fig. 2. $\tau(r; \Delta; \mu) = \tilde{K}^r_{0, \Delta + r} / K^r_{0, \Delta + r}$ vs. r for various Δ and $\mu = 0$.
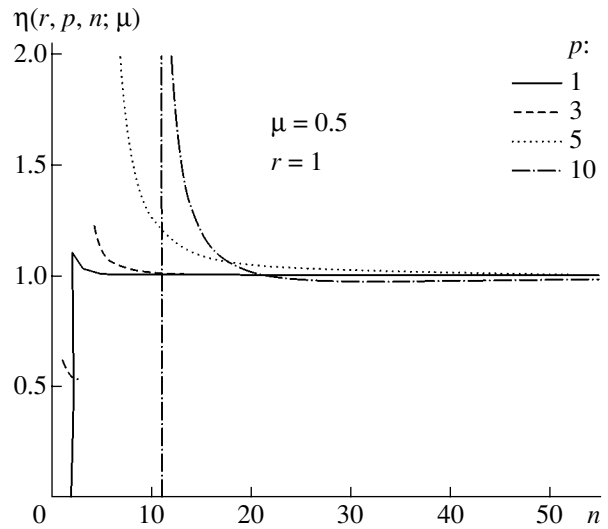
$\eta(r, p, n; \mu)$

Fig. 3. $\eta(r, p, n; \mu) = \tilde{K}^r_{p, n} / K^r_{p, n}$ vs. n for nonlinear MEs at various p.

Dependence (56) has a singularity at $\Theta = 0$. When calculating J_q and $\tilde{J}_q$ by formula (13), we used the approximation

$$F_{\mathrm{M}}(\Theta) \approx F_{\mathrm{M}}(z)$$

$$= \frac{\sqrt{1.5\pi}}{64 z^{5/4}}(1.00005 + 0.927579\sqrt{z} + 0.216338 z), \quad (57)$$

$$z = \sin^2 \Theta / 2.$$

With such an approximation, the error in determining $F_{\mathrm{M}}(\Theta)$ does not exceed 10^{-4}. Basically, the dependence $\eta(r; p, n; \mu) = \tilde{K}^r_{0, n} / K^r_{0, n}$ on the parameter $n(r)$ at $r \longrightarrow \infty$ ($n \longrightarrow \infty$) for Maxwellian molecules is the same as for pseudomolecules.

Figures 3 and 4 compare the asymptotics of the nonlinear MEs $K^r_{p, n}$ (see formulas (43), (47), and (53)) with the values found numerically for the pseudopotential at $\mu = 0.5$. In the former figure, the ratio $\eta(r, p, n; \mu) = \tilde{K}^r_{p, n} / K^r_{p, n}$ is plotted against n for fixed r and p. As for linear MEs, the place where the curve begins to saturate in terms of n depends considerably on the two other indices. The same situation takes place for the asymptotics in r, as follows from Fig. 4. Such behavior is explained by nonuniform asymptotic expansions (43), (47), and (53). The ratio $\tau(r, p, \Delta; \mu) = \tilde{K}^r_{0, \Delta + r} / K^r_{p, \Delta + r}$ as a function of r with fixed p and Δ is shown in Fig. 5. The asymptotic value, in this case, is reached faster than for the asymptotics in both r and n.

ASYMPTOTICS OF MATRIX ELEMENTS FOR MAXWELLIAN MOLECULES

The asymptotic formulas derived in the previous sections refer to the cases where one or two ME indices tend to infinity. Of interest also is the asymptotics when all three indices are large. We will show how such asymptotic formulas can be found in the general case for Maxwellian molecules. To be more specific, we will consider molecules of the Maxwellian type, for which the cross section varies in inverse proportion to g and the angular dependence $F(z)$, through which the coefficients J_q and $\tilde{J}_q$ in (13) are determined, is arbitrary. For such molecules, the situation is the simplest, since MEs are other than zero only if the sum of subscripts equals a superscript.

Here, it is more convenient to normalize the angular part of the cross section as

$$F^*(z) = 4\pi F(z). \quad (58)$$

Note that, for truly Maxwellian molecules, the angular component $F^*(z)$ of the cross section has a nonintegrable singularity and is defined by the function $F_{\mathrm{M}}(z)$ given above.

Previously [4], the formula that relates nonlinear MEs through linear ones for Maxwellian molecules was derived:

$$K^r_{n, p} = \frac{r!}{p! n!} \sum_{k=0}^{p} (-1)^k \binom{p}{k} K^{k+n}_{k+n, 0}, \quad (59)$$

where $K^r_{p, 0}$ is given by formulas (10), (11), and (13).

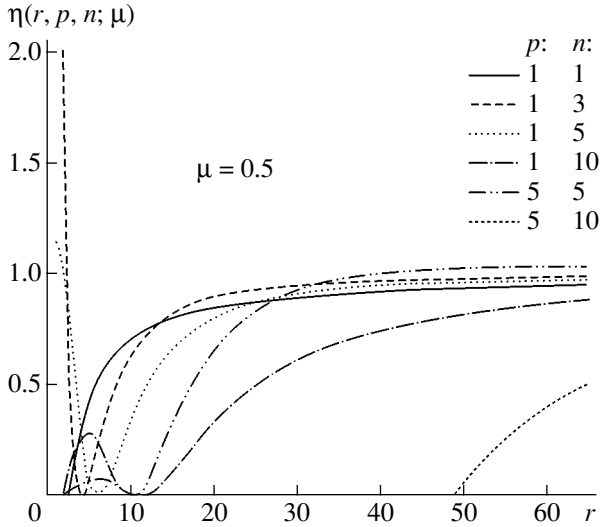For Maxwellian molecules, the expression under the summation sign in (10) is always zero except for the

$\eta(r, p, n; \mu)$

Fig. 4. $\eta(r, p, n; \mu) = \tilde{K}^r_{p,n}/K^r_{p,n}$ vs. r for nonlinear MEs at various p and n.
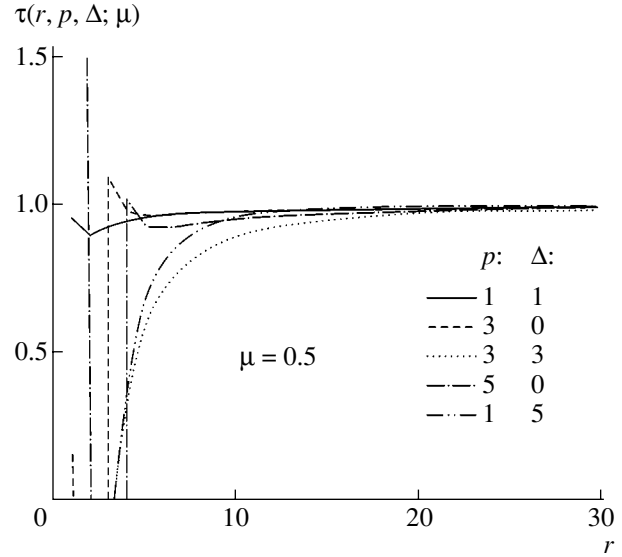
$\tau(r, p, \Delta; \mu)$

Fig. 5. $\tau(r, p, \Delta; \mu) = \tilde{K}^r_{p,\Delta+r}/K^r_{p,\Delta+r}$ vs. r for nonlinear MEs at various p and Δ.

case $n + r = 2q$. Since $q \leq \min(n, r)$, such an equality may hold only if $q = r = n$. Eventually, we have

$$K^r_{r,0} = \int_0^1 F^*(z)((1-z)^r - 1)dz. \quad (60)$$

Substituting (60) into (59) and changing the order of integration, we find

$$K^r_{n,p} = \frac{r!}{n!p!}\int_0^1 F^*(z)(1-z)^n \sum_{k=0}^p (-1)^k \binom{p}{k}(1-z)^k dz. \quad (61)$$

The sum in (61) is easy to calculate, and we arrive at

$$K^r_{n,p} = \frac{r!}{n!p!}\int_0^1 F^*(z)z^p((1-z)^{r-p} - \delta_{p,0}\delta_{r,n})dz. \quad (62)$$

In what follows, we will be interested only in nonlinear MEs; therefore, the last parenthesized term of the integrand may be omitted. In (62), it is more convenient to integrate over a symmetric interval by introducing a new variable of integration $z = (1-x)/2$:

$$K^r_{n,p} = \frac{r!}{2n!p!}\int_{-1}^1 \chi(x)\left(\frac{1+x}{2}\right)^n\left(\frac{1-x}{2}\right)^p dx. \quad (63)$$

Here, $\chi(x) = F^*((1-x)/2)$ and $x = \cos\Theta$.

Consider a function normalized within the interval $(-1, 1)$:

$$S_{n,N}(x) = \frac{1}{P}\left(\frac{1+x}{2}\right)^n\left(\frac{1-x}{2}\right)^{N-n}, \quad N = n + p,$$

$$P = 2\frac{n!(N-n)!}{(N+1)!}. \quad (64)$$

Substituting (64) into (63) and taking into account that $r = N$ for Maxwellian molecules, we come to

$$K^N_{n,p} = \frac{1}{N+1}\int_{-1}^1 \chi(x)S_{n,N}(x)dx. \quad (65)$$

It is easy to check that the function $S_{n,N}(x)$ has a single maximum at $x_0 = 2n/N - 1$. It is essential that the position of the maximum depends on the ratio n/N. It should be emphasized that nonlinear MEs have so far been expressed through the coefficients J_q and $\tilde{J}_q$, that is, through the moments of the scattering cross section angular component that have the functions $S_{0,N}(x)$ and $S_{N,N}(x)$ with peaks at the extremes of the interval. This is the reason why large-index nonlinear MEs are difficult to calculate. The use of moments with the function $S_{n,N}(x)$, for which the maximum position depends on n, eliminates this difficulty.

Figure 6 shows the function $S_{n,N}(x)$ for $N = 10$ and $n = 0, 2, 5$, and 10. It has a wide maximum, being symmetric at $n/N = 0.5$. As the quantity $2n/N - 1$ approaches the extremes of the interval, the function becomes more asymmetric.

Figure 7 shows the function $S_{n,N}(x)$ for different N and $n/N = 0.5$. The peak grows with N, becoming narrower. In the limit $N \longrightarrow \infty$, $S_{N/2,N}(x)$ tends to $\delta(x)$. For arbitrary n/N, the function $S_{n,N}(x)$ passes to $\delta(x - x_0)$ in the limit $N \longrightarrow \infty$. Now it becomes obvious that formula (65) asymptotically (at large N) tends to

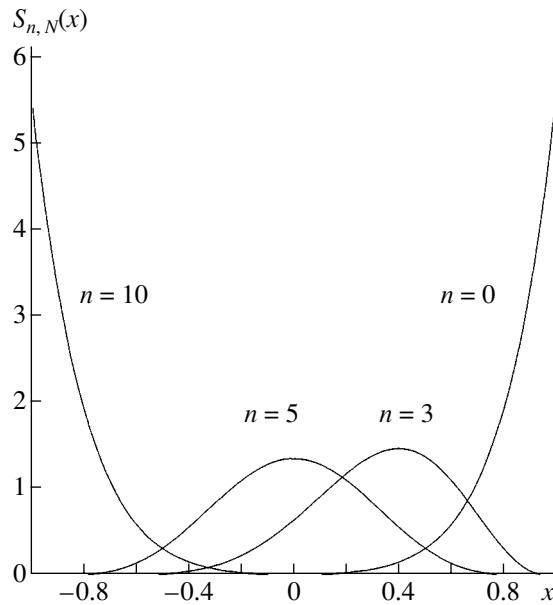$$K^N_{n,N-n} = \frac{1}{N+1}\chi(x_0) = \frac{1}{N+1}F^*(1 - n/N). \quad (66)$$

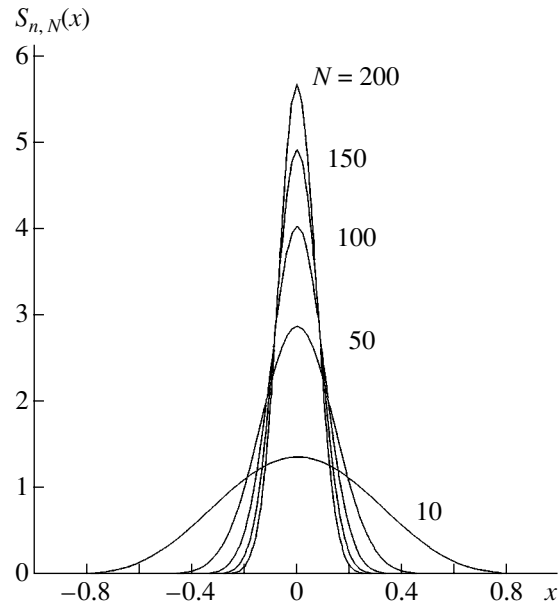**Fig. 6.** Function $S_{n,\,N}(x)$ at $N = 10$ for various n.

Fig. 7. Function $S_{n,\,N}(x)$ for various N and $n/N = 0.5$.

Figure 8 compares the quantity $(N + 1)K^N_{n,\,N-n}$ with the function $F^*(1 - n/N)$ for various N. Even with $N = 30$, the zeroth-order approximation is seen to give a good estimate of MEs. A more accurate result can be obtained from the asymptotic expansion of the integral in (65) for larger N.

Consider the integral

$$I_{p,\,n}(N) = \int_{-1}^{1} \chi(x) S_{n,\,N}(x) dx = \frac{1}{P} \int_{-1}^{1} \chi(x) \omega_{p,\,n}(x) dx. \quad (67)$$

Let the ratios $\alpha = p/N$ and $\beta = n/N$ remain constant at $N \longrightarrow \infty$. We represent the kernel $\omega_{p,\,n}(x)$ as

$$\omega_{p,\,n}(x) = \left(\frac{1+x}{2}\right)^n \left(\frac{1-x}{2}\right)^p = \exp\left(n \ln\left(\frac{1+x}{2}\right)\right.$$
$$\left. + p \ln\left(\frac{1-x}{2}\right)\right) = \exp(g(x)). \quad (68)$$

The function $g(x)$ has a maximum at the point $x_0 = \beta - \alpha$, which is naturally coincident with the position of the maximum of $\omega_{p,\,n}(x)$. The width of the maximum is found from relationship

$$|g''(x_0)|^{-1/2} = \frac{(4np)^{1/2}}{N^{3/2}} = \left(\frac{4\alpha\beta}{N}\right)^{1/2}.$$

Separating out the parameter N in the exponent of (68), we find for integral (67)

$$I_{p,\,n}(N) = \frac{1}{P} \int_{-1}^{1} \chi(x) \exp(Nf(x)) dx. \quad (69)$$

Here, the function

$$f(x) = \beta \ln\left(\frac{1+x}{2}\right) + \alpha \ln\left(\frac{1-x}{2}\right)$$

does not depend on N, and the integral itself can be estimated by the Laplace method. With $\alpha \neq 0$ and $\beta \neq 0$, the position of the peak lies within the segment $[-1, 1]$ and the asymptotics can be found by putting $x = \psi(\tau)$, where the function $\psi(\tau)$ is implicitly defined by the relationship $f(x_0) - f(x) = \tau^2$ and by the expansion of the load function $\chi(\psi(\tau))\psi'(\tau)$ into the series $\sum_{k=0}^{\infty} c_k \tau^k$ in the vicinity of the point $\tau = 0$ (see, e.g., [7]). This yields

$$I_{p,\,n}(N) = \frac{1}{P} e^{Nf(x_0)} \sqrt{\frac{\pi}{N}} \sum_{kk=0}^{\infty} \frac{c_{2k}(2k)!}{N^k \, k^k k!}$$
$$= \frac{1}{P} e^{Nf(x_0)} \sqrt{\frac{\pi}{N}} \left(c_0 + \frac{1}{2}\frac{c_2}{N} + O(N^{-2})\right). \quad (70)$$

The quantity P is the integral of the kernel $\omega_{p,\,n}(x)$ and can also be considered as an integral of Laplace type. The asymptotic estimate of the integral has the form

$$P = \int_{-1}^{1} 1 \omega_{n,\,p}(x) dx = \int_{-1}^{1} 1 e^{Nf(x)} dx$$

$$= e^{Nf(x_0)} \sqrt{\frac{\pi}{N}} \sum_{kk=0}^{\infty} \frac{d_{2k}(2k)!}{N^k \, 4^k k!} \quad (71)$$

$$= e^{Nf(x_0)} \sqrt{\frac{\pi}{N}} \left(d_0 + \frac{1}{2}\frac{d_2}{N} + O(N^{-2})\right),$$

where the coefficients d_k are found by expanding the function $\psi'(\tau)$ into a series.

Substituting (71) into (70) and calculating c_k and d_k yields

$$I_{p,n} = \chi(x_0) + 2[(\alpha - \beta)\chi'(x_0) + \alpha\beta\chi''(x_0)]\frac{1}{N}$$

$$+ \frac{2}{3}(6(\beta - \alpha)\chi'(x_0) + 3(2 - 11\alpha\beta)x''(x_0) \quad (72)$$

$$- 10(\beta - \alpha)\alpha\beta\chi'''(x_0) + 3\chi''''(x_0)(\alpha\beta)^2)\frac{1}{N^2} + O(N^{-3}).$$

Note that asymptotic expansion (72) makes sense even if $\alpha = p/N$ or $\beta = n/N$ vanishes. However, in this case, the position of the maximum is at an extreme of the segment and such a way of finding the asymptotics becomes to some extent invalid. To find the asymptotics of integral (67) at small β, we represent it in the form

$$I_{p,n}(N) = \frac{1}{P}\int_{-1}^{1} \chi(x)\left(\frac{1+x}{1-x}\right)^n\left(\frac{1-x}{2}\right)^N dx$$

$$= \frac{1}{P}\int_{-1}^{1} \chi(x)\left(\frac{1+x}{1-x}\right)^n \exp\left(N\ln\left(\frac{1-x}{2}\right)\right)dx. \quad (73)$$
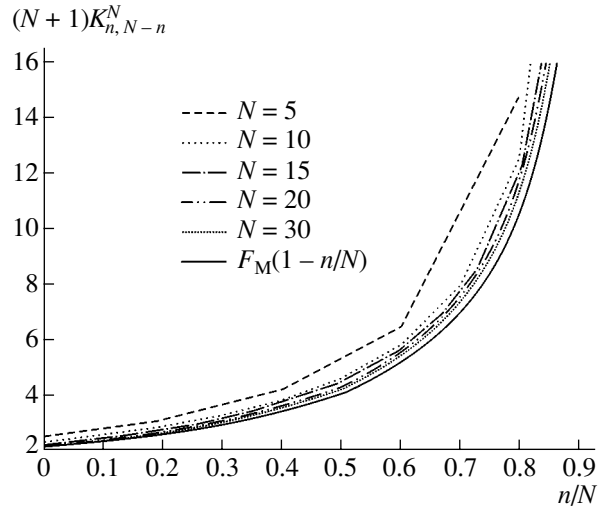
Changing the variable and expanding the load func-

$(N+1)K^N_{n,N-n}$

Fig. 8. $(N+1)K^N_{n,N-n}$ vs. n/M at various N. The function $F_M(1 - n/N)$ is plotted for comparison.

tion in the vicinity of the point $x = -1$, we find

$$I_{p,n}(N) = \chi(-1) + 2(n+1)\chi'(-1)\frac{1}{N} + (-4\chi'(-1)$$

$$+ 2(n+2)\chi''(-1))\frac{(n+1)}{N^2} + O(N^{-3}). \quad (74)$$

Note that expansion (74) coincides with (72) at $\beta = 0$. Hence, formula (72) can be used to asymptotically evaluate the MEs $K^N_{n,N-n}$ for any value of the index n.

Table 1. Zeroth- ($A0$), first- ($A1$), and second-order ($A2$) approximations of the quantity $(N+1)K^N_{(1-\alpha)N,\alpha N}$ calculated by asymptotic formula (72), as well as its accurate value (C) obtained by numerical calculation, for truly Maxwellian molecules for several $\alpha = p/N$

		$\alpha = 1/5$	$\alpha = 2/5$	$\alpha = 3/5$	$\alpha = 4/5$
$N = 5$	$A0$	2.647092	3.500223	5.259962	10.902165
	$A1$	3.157573	4.209428	6.423971	13.930613
	$A2$	3.166976	4.233389	6.503574	14.523547
	C	3.169085	4.238322	6.522939	14.790084
$N = 10$	$A0$	2.647092	3.500223	5.259962	10.902165
	$A1$	2.902332	3.854825	5.841966	12.416389
	$A2$	2.904683	3.860815	5.861867	12.564622
	C	2.904944	3.861414	5.864154	12.594127
$N = 20$	$A0$	2.647092	3.500223	5.259962	10.902165
	$A1$	2.774712	3.677524	5.550964	11.659277
	$A2$	2.775300	3.679021	5.555939	11.696335
	C	2.775333	3.679095	5.556218	11.699827
$N = 30$	$A0$	2.647092	3.500223	5.259962	10.902165
	$A1$	2.732172	3.618423	5.453963	11.406906
	$A2$	2.732433	3.619089	5.456175	11.423377
	C	2.732443	3.619112	5.455556	11.424028

Table 2. Zeroth- (*A*0), first- (*A*1), and second-order (*A*2) approximations of the quantity $(N + 1)K_{n, N-n}^{N}$ calculated by asymptotic formula (72), as well as its accurate value (*C*) obtained numerically for Maxwellian molecules with *n* close to *N*

		$n = N - 3$	$n = N - 2$	$n = N - 1$
$N = 5$	A0	3.500222	5.259962	10.902165
	A1	4.209428	6.423971	13.930612
	A2	4.233389	6.503574	14.523547
	C	4.238322	6.522939	14.790084
$N = 10$	A0	7.084825	10.902165	23.384551
	A1	7.934548	12.416389	28.036598
	A2	7.980223	12.564622	29.206591
	C	7.986631	12.594127	29.682732
$N = 20$	A0	14.908958	23.384551	51.526962
	A1	16.089988	25.710575	59.961619
	A2	16.176542	26.003073	62.438814
	C	16.187143	26.056194	63.389937
$N = 30$	A0	23.384551	37.017069	82.610505
	A1	24.935234	40.252079	95.319350
	A2	25.065233	40.702318	99.258378
	C	25.080483	40.781345	100.736674

Table 1 lists the zeroth- (*A*0), first- (*A*1), and second-order (*A*2) approximations of the quantity $(N + 1)K_{\beta N, \alpha N}^{N}$, as well as its accurate value (*C*) obtained numerically for Maxwellian molecules. The first approximation provides a good estimate of MEs even for $N = 5$; the second approximation gives the value of $(N + 1)K_{\beta N, \alpha N}^{N}$ accurate to fractions of a percent.

For Maxwellian molecules, the angular dependence $\chi(x)$ tends to infinity at $x \longrightarrow 1$, the derivatives increasing with their order. Therefore, in this range, the value of $(N + 1)K_{n, N-n}^{N}$ grows with *n* that is close to *N*, as well as with $N \longrightarrow \infty$, and the asymptotic estimate is found under very "stringent" conditions. However, approximation (72) remains good, as demonstrated by Table 2.

REFERENCES

1. G. Turchetti and M. Paolilli, Phys. Lett. A **90**, 123 (1982).
2. A. Ya. Énder and I. A. Énder, Zh. Tekh. Fiz. **64** (10), 38 (1994) [Tech. Phys. **39**, 997 (1994)].
3. A. Ya. Énder and I. A. Énder, Phys. Fluids **9**, 2720 (1999).
4. A. Ya. Énder and I. A. Énder, Zh. Tekh. Fiz. **72** (5), 1 (2002) [Tech. Phys. **47**, 513 (2002)].
5. H. Bateman and A. Erdlyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953; Nauka, Moscow, 1965).
6. G. E. Uhlenbeck and G. W. Ford, *Lectures in Statistical Mechanics* (American Mathematical Society, Providence, 1963; Mir, Moscow, 1965).
7. M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of Complex Variable* (Nauka, Moscow, 1973).

Translated by V. Isaakyan

THEORETICAL AND MATHEMATICAL PHYSICS

On the Theory of Thermodiffusiophoresis of a Moderately Coarse Two-Layer Volatile Aerosol Particle

S. N. D'yakonov and É. V. Efremov

Orel State University, Orel, 302015 Russia

Received September 13, 2002; in final form, March 6, 2003

Abstract—A hydrodynamic method with slip is used to construct a theory of uniform thermodiffusiophoretic transfer of a volatile high-viscosity two-layer particle under conditions in which one component of a moderately rarefied binary gas mixture undergoes phase transition on the particle's surface. The solid spherical core and high-viscosity sheath of the particle are concentric to each other. The theory suggests that, when moving, a two-layer volatile aerosol particle may be considered as a homogeneous particle with an effective thermal conductivity. The effect of the evaporation rate, interfacial temperature steps, and presence of the core on the rate of thermodiffusiophoresis is considered. Thermal diffusion terms, Stefan effects, and the heat evolution due to the convective transfer of the evaporating mass are taken into account. Under the same initial assumptions, formulas derived in this work are of wider practical significance than those following from the conventional approaches. © 2003 MAIK "Nauka/Interperiodica".

INTRODUCTION

Two-layer aerosol particles occur upon vapor condensation on solid centers and are frequently encountered in clouds. They are of special interest in applications concerned with environmental protection, in studying the capture (washing-out) of volatile (nonvolatile) or radioactive highly dispersed aerosols by coarser evaporating drops, in investigating aerosol deposition in condensing channels kept at different temperatures, in refining chemical engineering design schemes, etc. In this work, we generalize our original approach [1, 2], which takes into account an external gradient of the relative concentration of one component in a moderately rarefied binary gas mixture, interfacial temperature steps, and the presence of a center (core) with differing thermal properties.

STATEMENT OF THE PROBLEM, BASIC EQUATIONS AND BOUNDARY CONDITIONS

Let a two-layer spherical particle the outer surface of which has a radius of curvature R be placed in an infinite immobile binary gas mixture with constant collinear gradients of the temperature T, $\mathbf{A}_T$, and the relative concentration C, $\mathbf{A}_C$, of one component. The highly viscous sheath of the particle is a pure volatile liquid with an evaporation coefficient α. The solid core of the particle and its sheath are concentric to each other. The gas mixture, the sheath, and the core of radius R_* have thermal conductivities κ, κ', and κ'', respectively. The desired rate of thermodiffusiophoresis $\mathbf{U}_{ph}$ is reached when the resulting action of all force involved disappears [1]. The problem is solved in the coordinate system with the origin placed at the center of the core. The

polar axis z is directed along the vector $\mathbf{A}_T = (\nabla T)_\infty$. The velocity of the center of mass of the environment relative to the particle is $\mathbf{U} = -\mathbf{U}_{ph}$. The stationary temperature distribution $T''(\mathbf{r})$ inside the core satisfies the Laplace equation. The assumptions and notation are the same as in [1] (see also the end of the text). The conditions at infinity and at the interfaces are as follows.

$$r \longrightarrow \infty: \mathbf{v} = U\mathbf{i}_z, \quad T = T_0 + A_T z, \quad C = C_0 + A_C z,$$

$$r = R: \mathbf{s} \cdot \mathbf{v} = \mathbf{s} \cdot (K'_{Tsl}\nabla T + K'_{Dsl}\nabla C),$$

$$\mathbf{n} \cdot \left(n_1\mathbf{v} - \frac{(n_1+n_2)^2 m_2}{\rho}D(\nabla C + K_{TD}\nabla \ln T) \right)$$
$$= \alpha v(n_1+n_2)(C_s - C),$$

$$\mathbf{n} \cdot \left(n_2\mathbf{v} + \frac{(n_1+n_2)^2 m_1}{\rho}D(\nabla C + K_{TD}\nabla \ln T) \right) = 0,$$

$$T = T' + \mathbf{n} \cdot (V_{TT}\nabla T + V_{TC}T'\nabla C),$$

$$\mathbf{n} \cdot (-\kappa\nabla T + \kappa'\nabla T') = -Lm_1\alpha v(n_1+n_2)(C_s - C),$$

$$r = R_*: T' = T'', \quad \mathbf{n} \cdot (-\kappa'\nabla T' + \kappa''\nabla T'') = 0,$$

$$F_z = 0,$$

$$\rho = m_1 n_1 + m_2 n_2, \quad C = \frac{n_1}{n_1 + n_2},$$

$$C_s = \frac{n_{1s}}{n_1 + n_2}, \quad v = \left(\frac{kT}{2\pi m_1} \right)^{1/2}.$$

Here, the normally directed flux of the first component of the gas mixture at the interface is represented as the

normally directed flow of the volatile vapor, which is removed from the interface through the Knudsen layer and is proportional to the evaporation coefficient $\alpha \geq 0$ [1–3]. The substitution of the fraction $2\alpha/(2 - \alpha)$ for the evaporation (condensation) coefficient α generalizes the Hertz–Knudsen classical condition [4–6].

However, throughout this work, we use the close approximation of this fraction, $\alpha \ll 1$ (numerical estimates show that the rate of the thermophorertic transfer of an aerosol particle is almost independent of this coefficient when $\alpha \sim 1$; at weak or moderate volatile evaporation, this dependence is significant [1]). Temperature steps are due to local gradients $\nabla_n T$ and $\nabla_n C$ in the Knudsen layer that are normal to the interface. The gas-kinetic coefficients V_{TT} and V_{TC} allow us to estimate the effect of this layer on the vector velocity field $\mathbf{v}(\mathbf{r})$; scalar distributions $C(\mathbf{r})$, $T(\mathbf{r})$, $T'(\mathbf{r})$, and $T''(\mathbf{r})$; and the desired rate U of the thermophoretic transfer of a moderately coarse two-layer particle. Both coefficients are of the order of $O(\lambda)$. The analytical form of the coefficients of steps in temperature and in relative concentration of a volatile component in a binary gas mixture for arbitrary ratios of the molecular masses and component concentrations, together with techniques for evaluating the coefficients, is given in [7–21]. The parameters K_{Tsl}, K_{Dsl}, V_{TT}, and V_{TC} do not influence the evaporation coefficient α and vice versa. Therefore, in the numerical analysis which follows, slip coefficients and interfacial temperature steps are assigned to a nonvolatile particle. On the core surface $r = R_*$, the temperature and normal thermal flux are continuous. In the expansion of the temperature dependence of the relative concentration of the volatile saturated vapor into the Taylor series near $T = T_{sur}$, the first two terms are left [1].

Introducing the dimensionless quantities

$$\tilde{\mathbf{r}} = \frac{\mathbf{r}}{R}, \quad \tilde{\mathbf{v}} = \frac{\mathbf{v}}{U}, \quad \tilde{F}_z = \frac{F_z}{6\pi\eta_0 R U},$$

$$\tilde{T} = \frac{T - T_0}{A_T R}, \quad \tilde{T'} = \frac{T' - T_0}{A_T R}, \tag{1}$$

$$\tilde{T''} = \frac{T'' - T_0}{A_T R}, \quad \tilde{C} = \frac{C - C_0}{A_C R},$$

and omitting the tilde, we may state the boundary conditions in the following linearized form:

$$r \longrightarrow \infty: \mathbf{v} = U\mathbf{i}_z, \quad T = z, \quad C = z; \tag{2}$$

$$r = 1: \left\{ C_0 + (1 - C_0)\frac{m_2}{m_1} \right\} U v_r = \alpha v$$

$$\times \left\{ C_s(\tau) + \frac{\partial C_s}{\partial T}\bigg|_{T = \tau} (T' - \tau) - C_0 - A_C R C \right\}; \tag{3}$$

$$(1 - C_0)\left\{ C_0 + (1 - C_0)\frac{m_2}{m_1} \right\} U v_r$$

$$+ D\left\{ A_C \frac{\partial C}{\partial r} + A_T \frac{K_{TD}}{T_0}\frac{\partial T}{\partial r} \right\} = 0; \tag{4}$$

$$U v_\theta = -K_{Tsl}\frac{\eta_0}{\rho_0 T_0} A_T \sqrt{1 - \xi^2}\frac{\partial T}{\partial \xi}$$

$$- K_{Dsl} D A_C \sqrt{1 - \xi^2}\frac{\partial C}{\partial \xi}; \tag{5}$$

$$\varepsilon_T T = \varepsilon_T T' + \varepsilon_T k_{TT}\frac{\partial T}{\partial r} + \varepsilon_C C_0 k_{TC}\frac{\partial C}{\partial r}; \tag{6}$$

$$-\frac{\kappa_0}{\kappa_0'}\frac{\partial T}{\partial r} + \frac{\partial T'}{\partial r} = -\frac{L m_1 \alpha v n_0}{A_T \kappa_0'}$$

$$\times \left\{ C_s(\tau) + \frac{\partial C_s}{\partial T}\bigg|_{T = \tau} (T' - \tau) - C_0 - A_C R C \right\}; \tag{7}$$

$$r = R_*: T' = T'', \quad -\frac{\kappa_0'}{\kappa_0''}\frac{\partial T'}{\partial r} + \frac{\partial T''}{\partial r} = 0; \tag{8}$$

$$F_z = 0. \tag{9}$$

Here, $-1 \leq \xi = \cos\Theta \leq +1$,

$$\eta_0 = \eta(T_0, C_0, p_0), \quad \kappa_0 = \kappa(T_0, C_0, p_0),$$

$$\kappa_0' = \kappa'(T_0, p_0), \quad \kappa_0'' = \kappa''(T_0, p_0),$$

$$\tau = \frac{T_{sur} - T_0}{A_T R} \ll 1, \quad \varepsilon_T = \frac{A_T R}{T_0} \ll 1, \quad \varepsilon_C = \frac{A_C R}{C_0} \ll 1,$$

$$k_{TT} = \frac{V_{TT}}{R} \sim O(\text{Kn}), \quad k_{TC} = \frac{V_{TC}}{R} \sim O(\text{Kn}).$$

RATE OF THERMODIFFUSIOPHORESIS

$$v_r(r, \xi) = P_1(\xi) - \sum_{n = 2}^{\infty} \{ B_n r^{-n - 1} + D_n r^{-n + 1} \} P_{n-1}(\xi),$$

$$v_\theta(r, \xi) = -2\frac{J_2(\xi)}{\sqrt{1 - \xi^2}} - \sum_{n = 2}^{\infty} \{ (n - 1) B_n r^{-n - 1}$$

$$+ (n - 3) D_n r^{-n + 1} \}\frac{J_n(\xi)}{\sqrt{1 - \xi^2}},$$

$$C(r, \xi) = r\xi + \sum_{n = 0}^{\infty} L_n r^{-n - 1} P_n(\xi),$$

$$T(r, \xi) = r\xi + \sum_{n=0}^{\infty} F_n r^{-n-1} P_n(\xi),$$

$$T'(r, \xi) = \sum_{n=0}^{\infty} \{G'_n r^n + H'_n r^{-n-1}\} P_n(\xi),$$

$$T''(r, \xi) = \sum_{n=0}^{\infty} G''_n r^n P_n(\xi).$$

The properties of Gegenbauer (ultraspherical) polynomials [1] are defined by the equations ($n \geq 2$)

$$C_s(\tau) - C_0 + \left.\frac{\partial C_s}{\partial T}\right|_{T=\tau} (G'_0 + H'_0 - \tau) - \varepsilon_C C_0 L_0 = 0, \quad (3a)$$

$$-\left\{C_0 + (1 - C_0)\frac{m_2}{m_1}\right\} U(-1 + B_2 + D_2)$$

$$= \alpha\nu\left\{\left.\frac{\partial C_s}{\partial T}\right|_{T=\tau} (G'_1 + H'_1) - \varepsilon_C C_0(1 + L_1)\right\}, \quad (3b)$$

$$-\left\{C_0 + (1 - C_0)\frac{m_2}{m_1}\right\} U(B_{n+1} + D_{n+1})$$

$$= \alpha\nu\left\{\left.\frac{\partial C_s}{\partial T}\right|_{T=\tau} (G'_n + H'_n) - \varepsilon_C C_0 L_n\right\}, \quad (3c)$$

$$\varepsilon_C C_0 L_0 + \varepsilon_T K_{TD} F_0 = 0, \quad (4a)$$

$$(1 - C_0)\left\{C_0 + (1 - C_0)\frac{m_2}{m_1}\right\} U(-1 + B_2 + D_2)$$

$$= \frac{D}{R}\{\varepsilon_C C_0(1 - 2L_1) + \varepsilon_T K_{TD}(1 - 2F_1)\}, \quad (4b)$$

$$(1 - C_0)\left\{C_0 + (1 - C_0)\frac{m_2}{m_1}\right\} U(B_{n+1} + D_{n+1})$$

$$+ \frac{D}{R}(n+1)\{\varepsilon_C C_0 L_n + \varepsilon_T K_{TD} F_n\} = 0, \quad (4c)$$

$$U(2 + B_2 - D_2) = 2K_{Tsl}\frac{\eta_0}{\rho_0 T_0} A_T(1 + F_1)$$

$$+ 2K_{Dsl}\frac{D}{R}\varepsilon_C C_0(1 + L_1), \quad (5a)$$

$$U\{n B_{n+1} + (n-2)D_{n+1}\}$$

$$= n(n+1)\left\{K_{Tsl}\frac{\eta_0}{\rho_0 T_0} A_T F_n + K_{Dsl}\frac{D}{R}\varepsilon_C C_0 L_n\right\}, \quad (5b)$$

$$\varepsilon_T(1 + k_{TT})F_0 - \varepsilon_T(G'_0 + H'_0) + \varepsilon_C C_0 k_{TC} L_0 = 0, \quad (6a)$$

$$\varepsilon_T(1 + 2k_{TT})(1 + F_1) - \varepsilon_T(G'_1 + H'_1)$$
$$+ 2\varepsilon_C C_0 k_{TC}(1 + L_1) = 3(\varepsilon_T k_{TT} + \varepsilon_C C_0 k_{TC}), \quad (6b)$$

$$\varepsilon_T\{1 + (n+1)k_{TT}\}F_n - \varepsilon_T(G'_n + H'_n)$$
$$+ (n+1)\varepsilon_C C_0 k_{TC} L_n = 0, \quad (6c)$$

$$\frac{\kappa_0}{\kappa'_0}F_0 - H'_0 = -\frac{L m_1 \alpha\nu n_0}{A_T \kappa'_0}$$

$$\times\left\{C_s(\tau) - C_0 + \left.\frac{\partial C_s}{\partial T}\right|_{T=\tau} (G'_0 + H'_0 - \tau) - \varepsilon_C C_0 L_0\right\}, \quad (7a)$$

$$\frac{\kappa_0}{\kappa'_0}(-1 + 2F_1) + G'_1 - 2H'_1 = -\frac{L_1 m_1 \alpha\nu n_0}{A_T \kappa'_0}$$

$$\times\left\{\left.\frac{\partial C_s}{\partial T}\right|_{T=\tau} (G'_1 + H'_1) - \varepsilon_C C_0(1 + L_1)\right\}, \quad (7b)$$

$$\frac{\kappa_0}{\kappa'_0}(n+1)F_n + nG'_n - (n+1)H'_n = -\frac{L_1 m_1 \alpha\nu n_0}{A_T \kappa'_0}$$

$$\times\left\{\left.\frac{\partial C_s}{\partial T}\right|_{T=\tau} (G'_n + H'_n) - \varepsilon_C C_0 L_n\right\}, \quad (7c)$$

$$R_* G'_0 + H'_0 = R_* G''_0, \quad H'_0 = 0, \quad (8a)$$

$$R_*^3 G'_1 + H'_1 = R_*^3 G''_1,$$

$$-\frac{\kappa'_0}{\kappa''_0}R_*^3 G'_1 + 2\frac{\kappa'_0}{\kappa''_0}H'_1 + R_*^3 G''_1 = 0, \quad (8b)$$

$$R_*^{2n+1} G'_n + H'_n = R_*^{2n+1} G''_n,$$

$$-\frac{\kappa'_0}{\kappa''_0}R_*^{2n+1} nG'_n + \frac{\kappa'_0}{\kappa''_0}(n+1)H'_n + R_*^{2n+1} nG''_n = 0. \quad (8c)$$

In (8b), we may put

$$\delta_* = \left(1 - \frac{\kappa'_0}{\kappa''_0}\right)\left(1 + 2\frac{\kappa'_0}{\kappa''_0}\right)^{-1} R_*^3.$$

The linear dimensions and thermal properties of the solid core spherical core and high-viscosity concentric sheath ($\kappa'_0 \neq \kappa''_0$) are defined by the single parameter

$$\kappa_* = \frac{1 + 2\delta_*}{1 - \delta_*}\kappa'_0 = \begin{cases} \kappa'_0, & \text{if } R_* = 0 \\ \kappa''_0, & \text{if } R_* = 1. \end{cases} \quad (10)$$

Then, upon thermodiffusiophoresis, such a spherical two-layer aerosol particle may be considered as a homogeneous object with an effective thermal conductivity κ_*.

The z projection of the velocity of the center of inertia of the gaseous environment relative to the particle is given by

$$U = 2K_{Ts1}\frac{\eta_0}{\rho_0 T_0}A_T\frac{\delta'}{\delta} + 2K_{Ds1}\frac{D}{R}\frac{\delta''}{\delta}$$

$$+ \frac{\alpha\nu}{C_0 + (1-C_0)\frac{m_2}{m_1}}\left\{2\frac{\kappa_0}{\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau} - \varepsilon_T K_{TD}\right.$$

$$\left. - \varepsilon_C C_0\left(1 + 2\frac{\kappa_0}{\kappa_*} + 2(k_{TT} - k_{TC}K_{TD})\right)\right\}\frac{1}{\delta}, \quad (11)$$

where

$$\delta = \left(1 + 2\frac{\kappa_0}{\kappa_*}\right)\left(2 + (1-C_0)\frac{\alpha\nu R}{D}\right)$$

$$+ 2\frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}\left(\frac{\partial C_s}{\partial T}\bigg|_{T=\tau} + \varepsilon_T K_{TD}\right)$$

$$+ 2k_{TT}\left\{1 - C_0\frac{\alpha\nu R}{D} + 2\left(1 + \frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right)\right\} \quad (12)$$

$$- 4\frac{k_{TC}}{\varepsilon_T}\left\{\frac{\kappa_0}{\kappa_*}(1-C_0)\frac{\alpha\nu R}{D}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right.$$

$$\left. + \varepsilon_T K_{TD}\left(1 + \frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right)\right\},$$

$$\delta' = \frac{\kappa_0}{\kappa_*}\left(2 + (1-C_0)\frac{\alpha\nu R}{D}\right) + (\varepsilon_T K_{TD} + \varepsilon_C C_0)\frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}$$

$$+ k_{TT}\left\{(1-C_0)\frac{\alpha\nu R}{D} + 2\left(1 + \frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right)\right\}$$

$$\quad (13)$$

$$- \frac{k_{TC}}{\varepsilon_T}\left\{\left(2\frac{\kappa_0}{\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau} - \varepsilon_C C_0\right)(1-C_0)\frac{\alpha\nu R}{D}\right.$$

$$\left. + 2\varepsilon_T K_{TD}\left(1 + \frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right)\right\},$$

$$\delta'' = \frac{\kappa_0}{\kappa_*}(1-C_0)\frac{\alpha\nu R}{D}\left(\frac{\partial C_s}{\partial T}\bigg|_{T=\tau} + 2\varepsilon_C C_0\right)$$

$$+ (\varepsilon_T K_{TD} + \varepsilon_C C_0)\left(1 + \frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right)$$

$$+ 2k_{TT}\varepsilon_C C_0\left(1 + \frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right) \quad (14)$$

$$- 2\frac{k_{TC}}{\varepsilon_T}\varepsilon_C C_0\left(\frac{\kappa_0}{\kappa_*}(1-C_0)\frac{\alpha\nu R}{D}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right.$$

$$\left. + \varepsilon_T K_{TD}\left(1 + \frac{Lm_1\alpha\nu n_0}{A_T\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right)\right).$$

The parameters δ, δ', and δ'' linearly depend on the evaporation coefficient α:

$$\frac{dU}{d\alpha} = -2\nu\Phi\left(\kappa_0, \kappa_*\right)\left\{2\frac{\kappa_0}{\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau}\right) - \varepsilon_T K_{TD}$$

$$- \varepsilon_C C_0\left(1 + 2\frac{\kappa_0}{\kappa_*} + 2(k_{TT} - k_{TC}K_{TD})\right)\right\}\frac{1}{\delta^2},$$

where

$$\Phi(\kappa_0, \kappa_*) = 2K_{Ts1}\frac{\eta_0}{\rho_0 D}\left(k_{TC}(1-C_0) + \frac{Lm_1 n_0 D}{T_0\kappa_*}\right)$$

$$- K_{Ds1}\left(\left(1 + 2\frac{\kappa_0}{\kappa_*} + 2k_{TT}\right)(1-C_0) + 2k_{TD}\frac{Lm_1 n_0 D}{T_0\kappa_*}\right)$$

$$- \frac{1}{C_0 + (1-C_0)\frac{m_2}{m_1}}\left(1 + 2\frac{\kappa_0}{\kappa_*} + 2(k_{TT} - k_{TC}K_{TD})\right).$$

ANALYSIS OF RESULTS

The derivative $dU/d\alpha$ changes sign twice when

$$\Phi(\kappa_0, \kappa_*) = 0,$$

$$2\frac{\kappa_0}{\kappa_*}\frac{\partial C_s}{\partial T}\bigg|_{T=\tau} - \varepsilon_T K_{TD}$$

$$- \varepsilon_C C_0\left(1 + 2\frac{\kappa_0}{\kappa_*} + 2(k_{TT} - k_{TC}K_{TD})\right) = 0.$$

These equalities are satisfied for spherical homogeneous bodies with a thermal conductivity $\kappa_0 \approx \kappa_*$ (low thermal conductivity) and $\kappa_0 \ll \kappa_*$ (high thermal conductivity), respectively. These conditions meet with the corresponding equations [1] for thermophoresis ($A_C = 0$) of a volatile homogeneous sphere ($R_* = 0$) without considering temperature jumps ($k_{TT} = k_{TC} = 0$).

Figure 1 plots κ_*/κ_0' against the reduced radius R_* of the spherical core at different ratios κ_0'/κ_0''. In general, when a spheroidal aerosol particle with a spherical

core evaporates, its outer radius R decreases and the reduced radius R_+ accordingly grows. As follows from Fig. 1, the effective thermal conductivity increases ($\kappa_0'/\kappa_0'' < 1$) or drops ($\kappa_0'/\kappa_0'' > 1$) and reaches κ_0'' (that is, $\kappa_* = \kappa_0''$) at $R_* = 1$.

The effect of volatility on the thermal diffusion scalar fields and rate of particle transfer may be neglected when

$$\frac{\alpha v R}{D} \ll 1, \quad C_s(T_0)\left(\frac{L\mu}{R_g T_0} - 1\right) \sim |K_{TD}|,$$

$$k_{TC}|K_{TD}| \sim 1, \quad \frac{Lm_1 \alpha v R n_0}{\kappa_*} \ll k_{TC}T_0.$$

The temperature distribution and the distribution of the component's relative concentration are virtually independent of the thermal conductivity of the gaseous environment if the thermal conductivity of the volatile particle is high. A gaseous mixture that forms around the particle transfers an uncompensated momentum to it largely by thermal diffusion:

$$\delta \longrightarrow (1 + 2k_{TT})(1 - C_0)\frac{\alpha v R}{D} + 2\varepsilon_T K_{TD}\frac{Lm_1 \alpha v n_0}{A_T \kappa_*}$$
$$+ 2(1 + 2(k_{TT} - k_{TC}K_{TD}))\left(1 + \frac{Lm_1 \alpha v n_0}{A_T \kappa_*}\frac{\partial C_s}{\partial T}\Big|_{T=\tau}\right),$$

$$\delta' \longrightarrow \left(k_{TT} + \varepsilon_C C_0\frac{k_{TC}}{\varepsilon_T}\right)(1 - C_0)\frac{\alpha v R}{D}$$
$$+ (\varepsilon_T K_{TD} + \varepsilon_C C_0)\frac{Lm_1 \alpha v n_0}{A_T \kappa_*}$$
$$+ 2(k_{TT} - k_{TC}K_{TD})\left(1 + \frac{Lm_1 \alpha v n_0}{A_T \kappa_*}\frac{\partial C_s}{\partial T}\Big|_{T=\tau}\right),$$

$$\delta'' \longrightarrow (\varepsilon_T K_{TD} + \varepsilon_C C_0)\left(1 + \frac{Lm_1 \alpha v n_0}{A_T \kappa_*}\frac{\partial C_s}{\partial T}\Big|_{T=\tau}\right)$$
$$+ 2\varepsilon_C C_0(k_{TT} - k_{TC}K_{TD})\left(1 + \frac{Lm_1 \alpha v n_0}{A_T \kappa_*}\frac{\partial C_s}{\partial T}\Big|_{T=\tau}\right).$$

Numerical estimates made for a coarse ($R = 100\ \mu m$) and a moderately coarse ($R = 10\ \mu m$) homogeneous drop of ethanol in a C_2H_5OH–N_2 binary gas mixture show that the rate of thermophoresis ($A_C = 0$) strongly depends on α if the evaporation rate of the drop is low. With $0.05 < \alpha < 1$, this dependence is extremely weak [1]. At near-room temperatures, the volatility of the alcohol increases the rate of thermophoretic transfer by 10–13% compared with the rate of thermophoresis for a solid nonvolatile particle ($C_0 = 0.001$–0.100, $k_{TC} = 0$, $K_{TD} = 0$). This conclusion is consistent with results obtained by Bakanov [22, 23], who considered the thermophoresis of a solid nonvolatile aerosol particle.
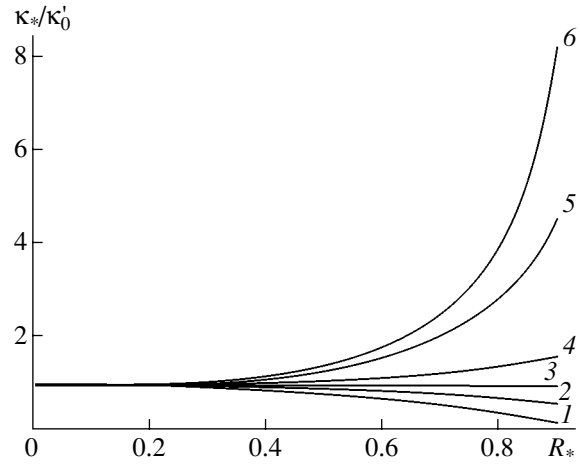

Fig. 1. κ_*/κ_0' vs. reduced radius R_*. $\kappa_0'/\kappa_0'' = $ (*1*) 100, (*2*) 2, (*3*) 1, (*4*) 0.5, (*5*) 0.1, and (*6*) 0.01.

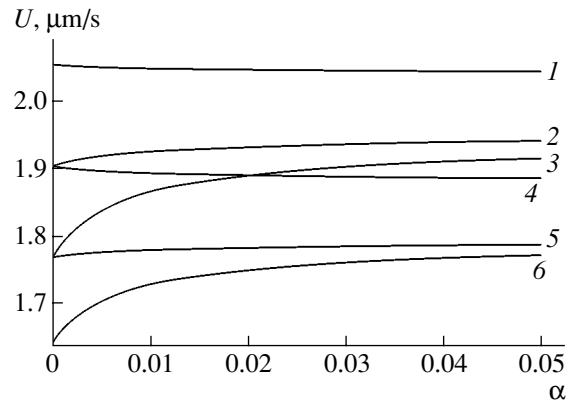**Fig. 2.** Rate U of the thermophoresis of the ethanol drop vs. evaporation coefficient α at $A_T = 100$ K/m, $R = 10\ \mu m$, $C_0 = 0.01$, and $K_{TD} = 0$ (thermal diffusion is absent). The temperature step k_{TT} in the C_2H_5OH–N_2 binary gas mixture (*1–3*) is and (*4–6*) is not taken into account. The undisturbed temperatures T_0 are (*1, 4*) 283, (*2, 5*) 303, and (*3, 6*) 323 K.

However, Bakanov's results virtually coincide with some of the findings reported in the references of books [7, 24].

In the case of moderately large particles ($R = 10\ \mu m$), a temperature step raises the rate of thermophoresis by 6–8%, as follows from Fig. 2.

Figures 3 and 4 plot the α dependences of the ratios of the thermophoretic ($A_C = 0$) and diffusiophoretic rate ($A_T = 0$) to the corresponding velocity of a solid nonvolatile particle with and without considering the temperature step k_{TT} for different temperatures.

Our work develops an approach that is alternative to the well-known conventional theories [7, 25, 26]. The results presented here coincide with the predictions of
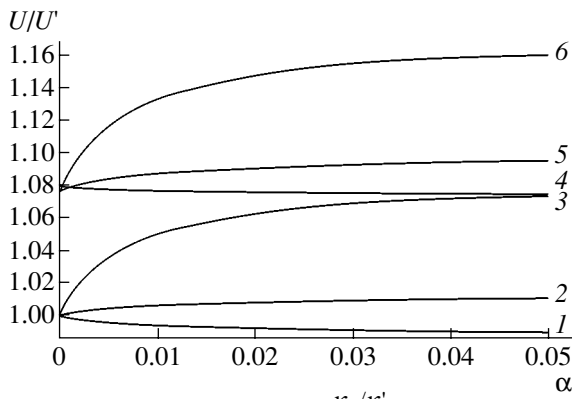
U/U'

Fig. 3. Ratio U/U' ($U' = 2K'_{Tsl}\dfrac{\kappa_0/\kappa'_0}{1 + 2\kappa_0/\kappa'_0} A_T$ is the rate of thermophoresis for a solid nonvolatile particle) as a function of evaporation coefficient α for $R = 10$ μm, $C_0 = 0.01$, and $K_{TD} = 0$ (thermal diffusion is absent). The temperature step k_{TT} (1–3) is not and (4–6) is taken into account. The undisturbed temperatures T_0 are (1, 4) 283, (2, 5) 303, and (3, 6) 323 K.
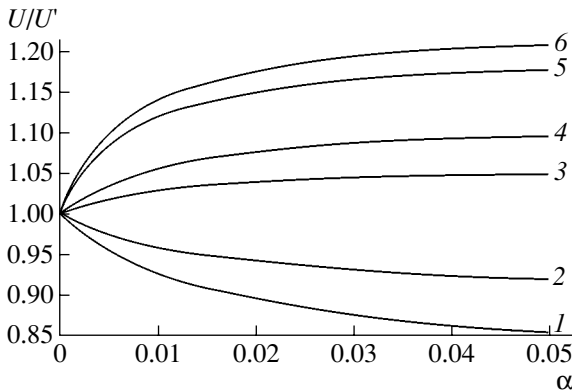
U/U'

Fig. 4. Ratio U/U' ($U' = K_{Dsl}DA_C$ is the rate of diffusiophoresis for a solid nonvolatile particle) as a function of evaporation coefficient α for $R = 10$ μm, $C_0 = 0.01$, and $K_{TD} = 0$ (thermal diffusion is absent). The temperature step k_{TT} (1, 3, 5) is not and (2, 4, 6) is taken into account. The undisturbed temperatures T_0 are (1, 2) 283, (3, 4) 303, and (5, 6) 323 K.
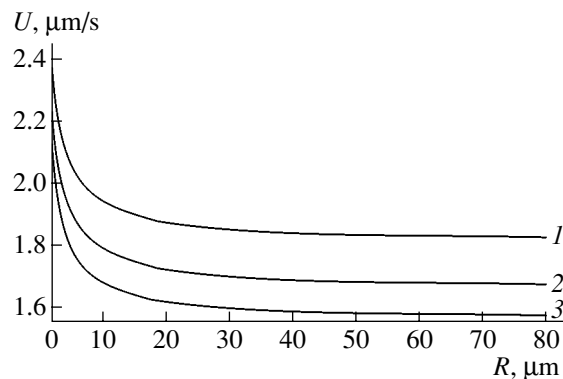
U, μm/s

Fig. 5. Transfer velocity U for a pure ethanol drop vs. its radius. (1) Formula from [25] and (2, 3) expressions (11)–(14) with $\alpha = 1$ and 0, respectively ($A_T = 100$ K/m, $T_0 = 323$ K, $C_0 = 0.1$, $k_{TC} = 0$, and $K_{TD} = 0$).

these theories only in the limiting case

$$C_0 \ll 1 \ll \frac{\alpha v R}{D}, \quad C_s(T_0) \ll 1, \quad \frac{n_0}{n'_0} \ll 1.$$

In the case of a moderately coarse volatile pure high-viscosity aerosol particle, the inequality $\alpha v R/D \gg 1$ fails even if the diffusion evaporation is high ($\alpha \sim 1$). When calculated from expressions (11)–(14) or by the formula derived in [25], the rate of thermophoretic transfer of a single moderately coarse volatile solid sphere has different values when the evaporation is weak and moderately high. As follows from Fig. 5, where the rate of thermophoresis U is plotted against the radius R of a spherical ethanol drop in the C_2H_5OH–N_2 gas mixture, the difference between our results and those of the conventional theory [25] reaches 15% when a coarse aerosol particle evaporates intensely. This difference increases to 30% if the binary mixture is saturated by the volatile component vapor, $C_0 \longrightarrow C_s(T_0)$, all other things being equal.

NOTATION

r, radius vector to the point of observation (m); (r, Θ, φ), spherical coordinates (m, rad, rad); $(\mathbf{i}_r, \mathbf{i}_\theta, \mathbf{i}_\varphi)$, corresponding dimensionless unit vectors; $(\mathbf{n}, \mathbf{s}, \mathbf{i}_\varphi)$, right-hand triple of dimensionless local characteristic unit vectors; (R, R_*), radii of curvature of the outer and inner surfaces of the spherical layer of the volatile high-viscosity sheath (m); **U**, velocity of the center of inertia of the binary gas mixture (m/s); $\mathbf{U}_{ph} = -\mathbf{U}$, velocity of uniform thermophoretic motion of the volatile particle (m/s); **F**, resultant force (N); Kn, dimensional Knudsen number; $\mathbf{v(r)}$, vector velocity field in the gaseous environment (m/s); $C(\mathbf{r})$, scalar field of a dimensional quantity—relative concentration of the first component in the gas mixture; $(T(\mathbf{r}), T'(\mathbf{r}), T''(\mathbf{r}))$, temperature distributions outside the two-layer particle, inside the sheath, and inside the core, respectively (K); T_{sur}, mean temperature on the surface of the aerosol particle (K); (p_0, T_0, C_0), undisturbed values of the pressure, temperature, and relative concentration of the volatile component (at the geometrical center of the core of the two-layer particle in its absence) (Pa, K, dimensionless); $(A_T = (\nabla T)_\infty, A_C = (\nabla C)_\infty$, constant gradients of the temperature and relative concentration of the volatile component in the gaseous environment at infinity (K/m, m^{-1}); (K_{Tsl}, K_{Dsl}), dimensionless gas-kinetic coefficients of thermal and diffusion slips of the binary gaseous environment; $(V_{TT}, V_{TC}$ and $k_{TT}, k_{TC})$, temperature steps at the boundary of the condensed phase (K, dimensionless); D, coefficient of interdiffusion of the gas mixture component (m^2/s); K_{TD}, dimensionless thermal diffusion ratio; $[(n_1, n_2), (m_1, m_2), (\lambda_1, \lambda_2)]$, numerical concentrations, masses, and mean free paths of gas molecules of the first and second sort (m^{-3}, kg, m); (ρ, η), density and dynamic viscosity of the gaseous environ-

ment (kg/m^3, (Pa s)); (κ, κ', κ''), thermal conductivities of the gaseous environment, sheath, and core, respectively (J/(K m)); (n_{1s}, C_s), absolute and relative concentrations of the saturated vapor of the condensed volatile phase (m^{-3}, dimensionless); L, heat of vaporization (J/kg); α, dimensionless evaporation coefficient; (μ, n'), molar mass and absolute concentration of molecules in the drop (kg/mol, m^{-3}); k, Boltzmann constant (J/K); R_g, gas constant (J/(K mol)); ∇, del operator (m^{-1}); E^2, second-order Stokes operator (m^{-2}); Δ is the Laplacian (m^{-2}); $\xi = \cos\Theta$, dimensionless angular variable; [$C_n^{-1/2}(\xi) = J_n(\xi)$, $C_n^{+1/2}(\xi) = P_n(\xi)$], Gegenbauer ultraspherical polynomials (dimensionless); (B, D, F, G', H', G'', L), unknown coefficients of natural-exponent expansions (dimensionless); (ε_T, ε_C), dimensionless small parameters; (δ_*, δ, δ', δ''), dimensionless parameters.

SUBSCRIPTS

Tsl, thermal slip; Dsl, diffusion slip; ph, phoretic motion; s, saturated vapor; g, gas; TD, thermal diffusion; α, reactive component; **v**, vector velocity field; T, temperature; (TT, TC) by temperature steps: the first T refers to temperature field, and the second indicates that the step is due to temperature (T) or concentration (C) discontinuity; 1, 2 by physical quantities refer to molecules of the first and second sort, respectively; 0 refers to undistorted parameters; z, projection onto the Oz axis; n, r, Θ mean projections onto the normal (**n**), radial (**r**), and tangential ($\mathbf{i}_\theta$) directions; sur, surface.

REFERENCES

1. S. N. D'yakonov, L. V. Kotlyarova, and Yu. I. Yalamov, Zh. Tekh. Fiz. **72** (3), 24 (2002) [Tech. Phys. **47**, 291 (2002)].

2. S. N. D'yakonov, É. V. Efremov, and A. A. Morozov, Zh. Tekh. Fiz. **72** (3), 11 (2002) [Tech. Phys. **47**, 306 (2002)].

3. Yu. I. Yalamov, On the Effect of Evaporation Coefficient on Diffusiophoresis of Coarse Particles, Available from VINITI, 1990, Moscow, no. 4120-B 90.

4. N. K. Makashev, Scientific Papers of Zhukovsky Central Institute of Aerohydrodynamics **5** (3), 49 (1974).

5. T. M. Muratova and D. A. Labuntsova, Teplofiz. Vys. Temp. **7**, 631 (1969).

6. N. V. Pavlyukevich, G. E. Gorelik, V. V. Levdanskiĭ, V. G. Leĭtsina, and G. I. Rudin, *Phase Transitions: Physical Kinetics and Transfer Processes*, Ed. by S. I. Anisimov (Nauka i Tekhnika, Minsk, 1980; Begell House, New York, 1995).

7. Yu. I. Yalamov and V. S. Galoyan, *Dynamics of Drops in Inhomogeneous Viscous Media* (Luĭs, Yerevan, 1985).

8. E. I. Alekhin and Yu. I. Yalamov, *Mathematical Grounds of Solving Boundary-Value Problems in the Kinetic Theory of Multicomponent Gases in the Vicinity of Condensed Phase* (Moskovsk. Pedagog. Univ., Moscow, 1991).

9. S. A. Savkov, A. A. Yushkanov, and Yu. I. Yalamov, Available from VINITI, 1986, Moscow, No. 5321.

10. Yu. I. Yalamov, A. A. Yushkanov, and S. A. Savkov, Dokl. Akad. Nauk SSSR **296**, 1107 (1987) [Sov. Phys. Dokl. **32**, 837 (1987)].

11. Yu. I. Yalamov, A. A. Yushkanov, and S. A. Savkov, Dokl. Akad. Nauk SSSR **301**, 1111 (1988) [Sov. Phys. Dokl. **33**, 615 (1988)].

12. M. N. Gaĭdukov, I. N. Ivchenko, and Yu. I. Yalamov, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 2, 199 (1972).

13. E. V. Metelkin and Yu. I. Yalamov, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 4, 142 (1973).

14. Yu. I. Yalamov and M. N. Gaĭdukov, Physics of Air-Dispersed Systems and Physical Kinetics (Kalinin, 1975), pp. 49–57.

15. Yu. I. Yalamov, E. R. Shchukin, and E. I. Alekhin, Topical Problems in Physics and Mechanics of Air-Dispersed Systems, Available from VINITI, 1989, Moscow, No. 580.

16. E. I. Alekhin and Yu. I. Yalamov, Selected Topics of Aerosol Physics, Available from VINITI, 1989, Moscow, No. 862, p. 3.

17. Yu. I. Yalamov, E. R. Shchukin, and E. I. Alekhin, Teplofiz. Vys. Temp. **28**, 256 (1990).

18. E. I. Alekhin and Yu. I. Yalamov, Kinetic Effects at the Interface between Liquid and Multicomponent Gas Mixture, Available from VINITI, 1990, Moscow, No. 4119.

19. Yu. I. Yalamov and A. A. Yushkanov, Physics of Air-Dispersed Systems and Physical Kinetics, Available from VINITI, 1979, Moscow, No. 3014, p. 149.

20. Yu. I. Yalamov, M. N. Gaĭdukov, V. S. Galoyan, and M. A. Melkumyan, Physics of Disperse Systems and Physical Kinetics, Available from VINITI, 1981, Moscow, No. 3865 (5), p. 7.

21. Yu. I. Yalamov, M. A. Melkumyan, and M. N. Gaĭdukov, Dokl. Akad. Nauk SSSR **270**, 1384 (1983) [Sov. Phys. Dokl. **28**, 499 (1983)].

22. S. P. Bakanov, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 181 (1995).

23. S. P. Bakanov, Kolloidn. Zh. **57**, 773 (1995).

24. E. R. Shchukin, Yu. I. Yalamov, and Z. L. Shulimanova, *Selected Problems in Physics of Aerosols* (Mosc. Pedagogical Univ., Moscow, 1992).

25. Yu. I. Yalamov, O. N. Zenkina, and M. F. Barinova, Inzh.-Fiz. Zh. **73**, 1295 (2000).

26. Yu. I. Yalamov and A. L. Lebedeva, Inzh.-Fiz. Zh. **73**, 1306 (2000).

Translated by V. Isaakyan

Effect of the Target Density on the Cross Section of Charge Exchange between Fast Ions and Atoms

O. Rosmej*, I. Yu. Tolstikhina**, and V. P. Shevelko**

* Gesellschaft für Schwerionenforschung (GSI), Darmstadt, Germany
*** Lebedev Physics Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia*
e-mail: shev@sci.lebedev.ru
Received September 4, 2002

Abstract—When fast X^{q^+} ions collide with atomic or molecular targets, the total charge exchange cross section decreases with increasing target density. This is because the excitation levels of resulting $X^{(q-1)^+}$ ions are suppressed because of ionization by target atoms. The effect of target density on the total charge exchange cross section may amount to one order of magnitude or more depending on the charge and energy of an incident ion, as well as on the density and inner shell configuration of target atoms. Numerical calculations are performed for partial (in the principal quantum number n) cross sections $\sigma(n)$ and total cross sections $\sigma_{\text{tot}} = \Sigma_n\sigma(n)$ of charge exchange in the case of collisions of fast multiply charged ions having an energy E in the range 100 keV/u–10 MeV/u with gas or solid targets. © 2003 MAIK "Nauka/Interperiodica".

INTRODUCTION

Charge exchange in atom–ion collisions,

$$X^{q^+} + A \longrightarrow [X^{(q-1)^+}]^* + A^+, \qquad (1)$$
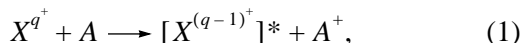
is an effective mechanism of generating excited ions $[X^{(q-1)^+}]^*$, where X^{q^+} is an incident ion with a charge q and A is a target atom. The distribution of ions $X^{(q-1)^+}$ over excited states, which depends on the cross sections of charge exchange passing an ion into certain quantum states (partial or selective charge exchange cross sections), defines the radiation of the ion after collision. The total (over all final states) charge exchange cross sections specify the lifetime and mean charge of ion beams when they interact with atoms and molecules of a residual gas in accelerators or with plasma targets [1, 2].

In recent years, the spectroscopic properties of resulting $X^{(q-1)^+}$ ions, especially their radiation in the X-ray and VUV spectral ranges, have attracted considerable attention from researchers engaged in the physics of plasma or those that study the interaction of ion beams with gaseous and solid targets [3–5]. This issue is also of interest to those researching the diagnostics of a laboratory plasma heated by neutral atomic beams [6, 7]. Therefore, detailed research into charge exchange, the distribution of resulting ions over excited states, and the determination of total cross sections are of undeniable interest.

In this study, we consider the effect of target density on the charge exchange cross section when fast multiply charged ions X^{q^+} collide with target atoms or molecules. It will be shown that, as the target density grows, the probability that $X^{(q-1)^+}$ ions will be brought into highly excited states decreases, since they are ionized by target atoms. As a result, the total charge exchange cross sections also decrease. The effect of target density on the total cross section may be relatively high: about one order of magnitude or higher.

Below, we use atomic units.

PARTIAL AND TOTAL CHARGE EXCHANGE CROSS SECTIONS

The cross sections of charge exchange that brings atoms (ions) into quantum states nlj (that is, partial cross sections),

$$X^{q^+} + A \longrightarrow [X^{(q-1)^+}(nlj)]^* + A^+ \qquad (2)$$
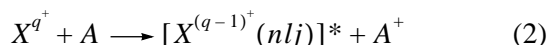
have certain distributions over the levels nlj depending on the collision conditions (here, n is the principal quantum number and l and j are the orbital and total moments of a captured electron in the final state). In the case of statistical distribution, the partial cross sections and total cross section can be represented as

$$\sigma_{nlj} = \frac{2j+1}{2(2l+1)}\sigma_{nl}, \quad \sigma_{nl} = \frac{2l+1}{n^2}\sigma_n, \qquad (3)$$

$$\sigma_{\text{tot}} = \sum_{nlj}\sigma_{nlj} = \sum_{n=n_0}^{\infty}\sigma_n, \qquad (4)$$

where n_0 is the ground state of the ion $X^{(q-1)^+}$ and σ_{nl} and σ_n are the cross sections averaged over the quantum numbers j and l, respectively.

In general, the distribution of ions $X^{(q-1)^+}(nlj)$ over levels nlj is defined by Stark mixing of these levels in an electric field that is produced in an ion beam or in a plasma by surrounding charged particles, and finding the distribution over nlj is a challenge. Experimental data for partial charge exchange cross sections are virtually absent: only total cross sections have been reported in the literature. However, in the case of multiply charged ions, the split of nlj levels may be considerably large. This allows one to measure the emission intensities from resulting ions, which are known to be sensitive to partial charge exchange cross sections. Associated experiments were carried out, e.g., in [8–10], where the partial cross sections of charge exchange between multiply charged ions and solid or gaseous targets were determined from the X-ray emission of related ions. It should be noted that such indirect measurements of cross sections face considerable computation difficulties associated with the need for calculating the energy levels and wavelengths, as well as the probabilities of radiative transitions in multiply charged ions.

When the collision energy is low, the total cross section of charge exchange between multiply charged ions and atoms is weakly dependent on energy and is given by [11]

$$\sigma_{\text{tot}} \approx \text{constan}\,t \approx \frac{q \times 10^{-15}}{(I_t/\text{Ry})^{3/2}}\text{cm}^2, \qquad (5)$$

where I_t is the binding energy of a target atom in the shell in Rydberg units (1 Ry = 13.606 eV).

The principal quantum number n_{max} of that state of the $X^{(q-1)^+}$ ion into which charge exchange brings the ion with the maximal probability is found from the classical model [12]:

$$n_{\text{max}} \approx q^{0.75}/(I_t/\text{Ry})^{0.5}. \qquad (6)$$

As the relative velocity v increases, the cross section of charge exchange from the shell with the binding energy I_t drops sharply by the law

$$\sigma \propto \frac{q^5 I_t^{5/2}}{v^{12}}, \quad v/q \gg 1 \qquad (7)$$

and has a distribution over principal quantum numbers n, which peaks at $n = n_{\text{max}}$ and falls as n^{-3} for $n \gg n_{\text{max}}$. As to the distribution over orbital quantum numbers l, ions in Rydberg states nl with $n \gg 1$ have, as a rule, $l = 0$, 1, or 2 if the collision energy is high (see, e.g., [10–12]). As the collision energy increases further, charge

exchange brings the resulting ion $X^{(q-1)^+}$ into the ground state $n_0 l_0$.

As is seen, the pattern of excited ion formation upon charge exchange with atoms is rather complicated even if the collision energy is high. For simplicity, we restrict the analysis to charge exchange between fast ions and atoms that brings the ions into states with certain principal quantum numbers n:

$$X^{q^+} + A \longrightarrow [X^{(q-1)^+}(n)]^* + A^+, \qquad (8)$$
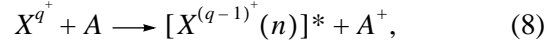
that is, the distributions of the resulting ions over the quantum numbers l and j will not be considered.

EFFECT OF TARGET DENSITY

The integral (including all states) charge exchange cross section depends not only on the relative velocity v and atomic configurations of colliding particles but also on the target density. For low-density targets, the summation over levels with quantum numbers n (see expression (4) for the total cross section) goes basically from the ground state $n = n_0$ to infinity, $n = \infty$. As the target density increases, the sum over n is cut off at a certain n_{cut}, since the states of $[X^{(q-1)^+}(n)]^*$ ions are suppressed by ionization collisions with target atoms:

$$[X^{(q-1)^+}(n)]^* + A \longrightarrow X^{q^+} + A + e^-. \qquad (9)$$
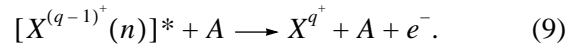
Thus, as the target density grows, the fraction of resulting ions in excited states decreases and so does the total charge exchange cross section.

Let us illustrate this using charge exchange between H-like oxygen ions and helium atoms as an example:

$$O^{7+} + \text{He} \longrightarrow O^{6+} + \text{He}^+. \qquad (10)$$
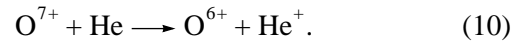
The effective cross sections of ionization, charge exchange, and excitation when O^{7+} oxygen atoms interact with He atoms at energies $E > 1$ keV/u are given in Fig. 1. The cross sections were computed with the CAPTURE [13] and LOSS [14] codes. In the energy range $E = 1$–100 keV/u, the exchange cross section, according to (5), is quasi-constant, $\sigma_{\text{es}} \approx 1.9 \times 10^{-15}$ cm², and then drops sharply by law (7). The cross section of O^{7+} ion ionization by He atoms reaches a maximum, $\sigma_{\text{ion}} \approx 3.0 \times 10^{-19}$ cm² at $E \approx 1$ MeV/u. Figure 1 also shows the excitation cross section for the $1s$–$2p$ transition in an O^{7+} ion colliding with a He atom. It is seen that charge exchange is the only process leading to the formation of excited O^{6+} ions for $E = 1$–800 keV/u.

The distribution of He-like oxygen ions arising after charge exchange over principal quantum numbers n is demonstrated in Fig. 2 for $E = 100$, 400, and 3200 keV/u. As the energy of incident ions grows, the distribution maximum shifts toward lower n. For $E = 3200$ keV/u, charge exchange brings the O^{6+} ion primarily into the ground $1s$ state ($n_0 = 1$).
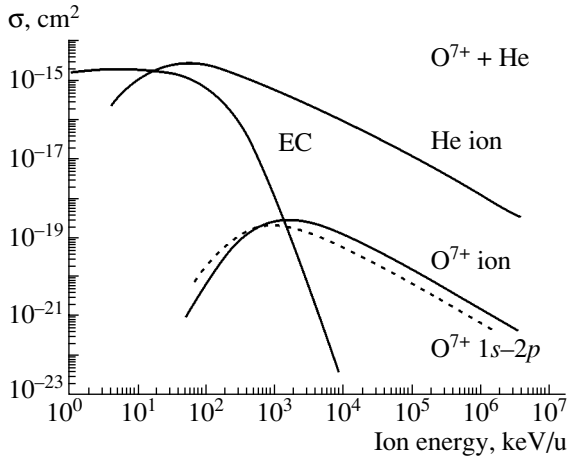
Fig. 1. Cross sections of charge exchange, ionization, and excitation upon collision of O^{7+} ions with He atoms. EC, charge exchange cross section; He ion, cross section of He ion ionization by O^{7+} ions; and O^{7+} ion, cross section of collision between O^{7+} ions and He atoms. The dashed line is the excitation cross section for the $1s–2p$ transition in an O^{7+} ion upon collision with a He atom (calculation by the CAPTURE and LOSS codes; this study).
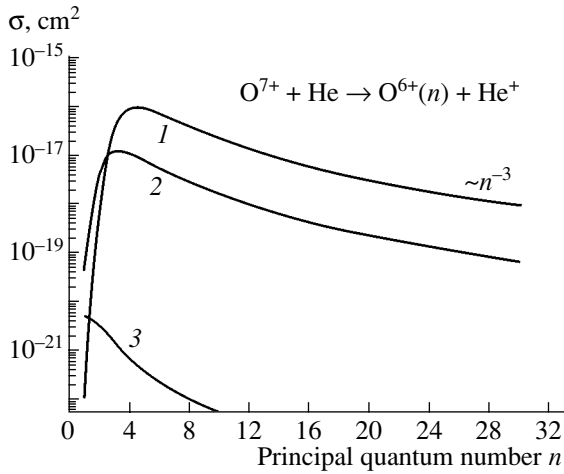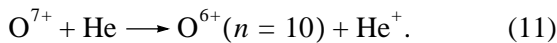

Fig. 2. Distribution of the cross sections of the charge exchange reaction $O^{7+} + He \longrightarrow O^{6+}(n) + He^+$ over the principal quantum numbers n for $E = (1)$ 100, (2) 400, and (3) 3200 keV/u. $\sigma_{tot} = (1)\ 6.0 \times 10^{-16}$, (2) 6.8×10^{-17}, and (3) 1.2×10^{-20} cm^2 (CAPTURE code; this study).

Let us make several numerical estimates for reaction (10) when the principal quantum number n equals 10:

$$O^{7+} + He \longrightarrow O^{6+}(n = 10) + He^+. \qquad (11)$$

The probability of the radiative decay of an excited O^{6+} ion $(n = 10)$ to all lower levels and the lifetime are estimated by the Kramers formula [15] as $A(n = 10) = 2.6 \times 10^9$ s^{-1} and $\tau(n = 10) = 3.8 \times 10^{-10}$ s. For this relatively long time, an excited O^{6+} ion $(n = 10)$ interacts

with He atoms, ionizing with a rate $\rho(He)v\sigma_{ion}$ s^{-1}, where $\rho(He)$ is the He atom density and σ_{ion} is the ionization cross section for O^{6+} ions in the state $n = 10$ at a relative velocity v. Note that the ionization cross sections for highly excited ($n \gg 1$) atoms and ions grow markedly with n: $\sigma_{ion} \sim n^2$. For example, with $E = 100$ keV/u ($v = 4.4 \times 10^8$ cm/s), σ_{ion} and $v\sigma_{ion}$ estimated by the Thomson formula [15] are, respectively, 8×10^{-16} cm^2 and 3.4×10^{-7} cm^3/s. The ionization rate $\rho(He)v\sigma_{ion}$ should be compared with the probability of radiative decay $A(n = 10) = 2.6 \times 10^9$ s^{-1}. It follows that, for the He atom density $\rho(He) > 2.6 \times 10^9$ s^{-1}/3.4×10^{-7} cm^3/s $\approx 7.5 \times 10^{15}$ cm^{-3}, all levels with quantum numbers $n > 10$ in the resulting O^{6+} ion will be ionized and make no contribution to the total charge exchange cross section.

Thus, with the target density taken into account, the total cross section takes the form

$$\sigma_{tot} = \sum_{n = n_0}^{n_{cut}} \sigma(n), \qquad (12)$$

where n_0 is the ground state of the $X^{(q-1)^+}$ ion and n_{cut} (cutoff parameter) is the upper limit of summation.

In general, the value of n_{cut} is estimated from the condition that the ionization rate equals the probability of radiative decay of the level n to all lower levels:

$$\rho_t \sigma_{ion}(n) v = A(n) = \sum_{n' \geq n_0}^{n-1} A_{nn'}, \qquad (13)$$

where ρ_t is the target atom density and $\sigma_{ion}(n)$ is the cross section of ionization of the atom $X^{(q-1)^+}$ from the state with a principal quantum state n at an ion velocity v.

Again using the Kramers formula for the total probability of decay $A(n)$ and the Thomson formula for the ionization cross section, we find from (13) an estimate of the maximal principal quantum number n_{cut} starting from which the levels of resulting ions decay by ionization due to collision with target atoms:

$$n_{cut} = n_0 + \Delta n,$$

$$\Delta n \approx q\left(\frac{10^{18}}{Z_t^2 \rho_t[cm^{-3}]}\right)^{1/7}\left(\frac{v^2}{10q^2}\right)^{1/14}$$

$$\approx q\left(\frac{10^{18}}{Z_t^2 \rho_t[cm^{-3}]}\right)^{1/7}\left(\frac{E[keV/u]}{250q^2}\right)^{1/14}, \qquad (14)$$

where ρ_t is the target atom density (cm^{-3}), v is the ion velocity in atomic units (1 a.u. $= 2.2 \times 10^8$ cm/s), and E is the energy of incident ions (keV/u).

From Eq. (14), it follows that, as the target becomes heavier and denser, more and more levels decay as a result of ionization; that is, charge exchange with excitation makes a lesser contribution to the total cross section. The value of n_{cut} depends on the relative velocity rather weakly: $n_{cut} \sim v^{1/7}$, as demonstrated in Table 1.

The expression for the ionization of atoms and ions by electrons that is similar to (14) was given in [15]. Unlike (14), n_{cut} in this expression depends on the electron temperature T_e in the plasma rather than on the ion beam energy.

NUMERICAL CALCULATIONS VERSUS EXPERIMENTAL DATA

In this study, the effect of target density on the charge exchange cross section is demonstrated with collisions of multiply charged ions with light atoms or multielectron targets (SiO_2) with a complex electronic configuration. The partial and total charge exchange cross sections were calculated by the CAPTURE code [13]. In brief, charge exchange cross sections are calculated as a function of the impact parameter with normalized charge exchange probabilities $W^N(b, v)$ in the form

$$\sigma_{01}(b, v) = 2\pi \int_0^\infty W^N_{01}(b, v)b\,db;$$

$$W^N_{01}(b, v) = \frac{W_{01}(b, v)}{1 + \sum_k W_{0k}(b, v)}. \qquad (15)$$

Here, b is the impact parameter, W^N is the charge exchange probability normalized to all possible channels k of $X^{(q-1)^+}$ ion formation, the probabilities $W^N \leq 1$, the probabilities W_{01} are calculated in the Brinkman–Kramers approximation with hydrogen-like wave functions (for details, see [13]), and the subscripts 0 and 1 refer to the initial and final states of the system.

Figure 3 shows the total cross section σ_{tot} of the charge exchange reaction

$$O^{7+} + He \longrightarrow O^{6+} + He^+$$
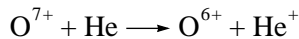
vs. He atom density with the energy $E = 100$ keV/u. In this specific case, as the He atom density ρ_t grows from 10^{19} to 10^{21} cm^{-3}, the total charge exchange cross section decreases by roughly one order of magnitude following the law $\sigma_{tot} \sim \rho_t^{-0.7}$. In general, σ_{tot} may vary with ρ_t in a different way.

For multielectron targets with a complex electronic configuration, the situation gets much more complicated because of the need for considering the capture of inner-shell electrons. This fact, the capture of inner-

Table 1. Limiting quantum numbers n_{cut} (condition (14)) that contribute to the total cross section of the charge exchange reaction $O^{7+} + He \longrightarrow O^{6+} + He^+$ vs. He atom density ρ_t and oxygen ion density E

ρ_t, cm^{-3}	$E = 100$ keV/u	$E = 400$ keV/u	$E = 3200$ keV/u
1×10^{10}	58	64	74
1×10^{13}	22	24	28
1×10^{16}	9	10	11
1×10^{19}	4	4	5
1×10^{22}	2	2	3

shell electrons, is most typical of charge exchange between fast ions and complex atoms (see, e.g., [11]). Figure 4 shows the contribution from inner-shell electrons of cesium (55 electrons) in the reaction of cesium–carbon charge exchange:

$$C^{6+} + Cs(1s^2 \ldots 5s^2 5p^6 6s) \longrightarrow C^{5+} + Cs^+. \qquad (16)$$
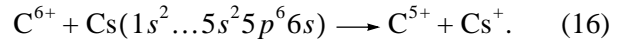
It is seen that, at $E > 30$ keV/u, charge exchange takes place only through the capture of inner-shell electrons of Cs. The capture of outer $6s$ electrons does not influence the process.

Figure 5 plots the cross section distribution over the principal quantum number n in the reaction of charge exchange between H-like Ge ions and Ne atoms,

$$Ge^{31+} + Ne \longrightarrow Ge^{30+}(n) + Ne^+ \qquad (17)$$
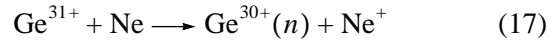
for the energy $E = 5.5$ MeV/u. As follows from the Monte Carlo calculations [16] and calculations performed in this work, charge exchange causes excited states with $n = 2$–5. The cross section of charge exchange into the ground state $n_0 = 1$ is three orders of magnitude smaller.
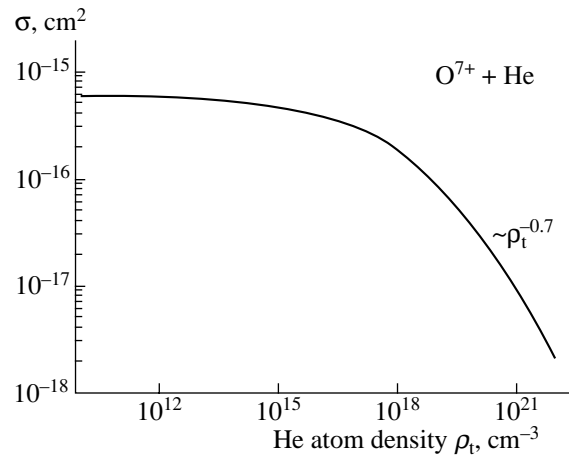

Fig. 3. Cross section of the charge exchange reaction $O^{7+} + He \longrightarrow O^{6+} + He^+$ vs. He atom density for $E = 100$ keV/u (CAPTURE code; this study).
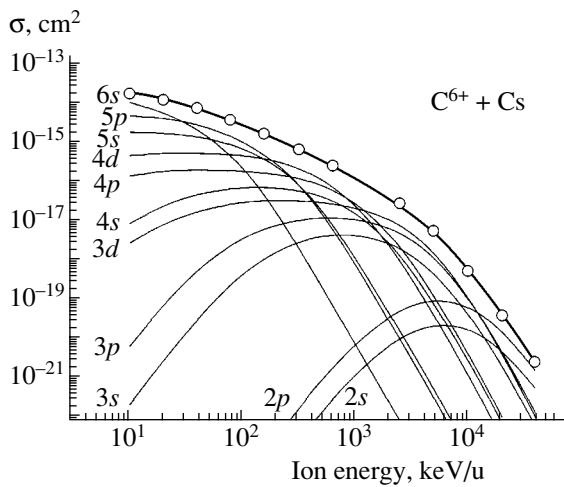
Fig. 4. Cross sections of the charge exchange between carbon ions and cesium atoms, $C^{6+} + Cs \longrightarrow C^{5+} + Cs^+$, vs. C^{6+} ion energy (low-density target). The contributions from different cesium atom subshells (continuous curves) to the total charge exchange cross section (the curve connecting the circles) are indicated (CAPTURE code; this study).

The total charge exchange cross sections for reaction (17) are presented in Fig. 6 for the energy interval $E = 3–12$ MeV/u. Here, the experimental data [17] are compared with the Monte Carlo calculations [16]; eikonal approximation [17]; and our calculations, where n_{cut} was varied from 15 to 8 and the energy was raised from 4.5 to 11.5 MeV/u, respectively.

The total charge exchange cross sections are often estimated by the Schlachter semiempirical formula [18]

$$\sigma_{tot}(E) = 1.1 \times 10^{-8} \text{ cm}^2 q^{0.5}/Z_t^{1.8} u^{-4.8}$$
$$\times [1 - \exp(-0.037u^{2.2})][1 - \exp(-2.44 \times 10^{-5} u^{2.6})], \quad (18)$$

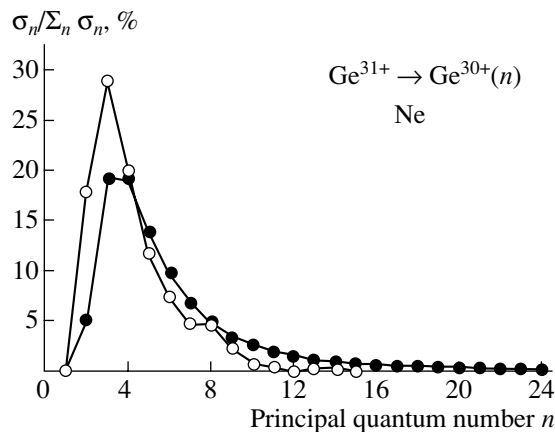$$u = E[\text{keV/u}]/(q^{0.7}Z_t^{1.25}), \quad u \geq 10, \quad q \geq 3,$$

Fig. 5. Calculated relative cross sections of charge exchange between H-like Ge ions and Ne atoms, $Ge^{31+} + Ne \longrightarrow Ge^{30+}(n) + Ne^+$, for an energy of 5.5 MeV/u: (○) Monte Carlo method and (●) CAPTURE code (this study).

where u is the scaled energy of a projectile ion and Z_t is the nuclear charge of a target atom.

At high energies, $u \gg 10$, cross section (18) has the asymptotics

$$\sigma_{tot}(E) \approx 1.1 \times 10^{-8} \text{ cm}^2 q^{3.9} Z_t^{4.2} E^{-4.8}[\text{keV/u}].$$

Schlachter formula (18) is correct up to a factor of 2–3 (similarly to the Lotz formulas for the ionization of atoms and ions by electron impact) and is very suitable for the estimation of charge exchange cross sections at medium and high energies. For low energies, $u < 1$, formula (18) does not apply. Schlachter estimates of the cross sections are also shown in Fig. 6.

Figures 7–10 show the total cross sections of charge exchange between fast nickel nuclei and SiO_2 molecules,

$$Ni^{28+} + SiO_2 \longrightarrow Ni^{27+} + [SiO_2]^+, \quad (19)$$
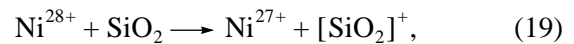
calculated at energies $E > 100$ keV/u and various SiO_2 densities (the state of SiO_2 was varied from gaseous to solid). The cross section of charge exchange with the molecules were calculated as the sum of the charge exchange cross sections per constituent atoms. At high energies, such an approach is a fairly close approximation.

Experimental X-ray investigation of heavy ion slowing-down in a material is currently proceeding at the GSI (Darmstadt, Germany) with the UNILAC accelerator [4]. In these experiments, targets are SiO_2 aerosols with a density varying over a wide range: from 0.04 g/cm³ ($\rho_t = 1.2 \times 10^{21}$ cm⁻³) to the density of
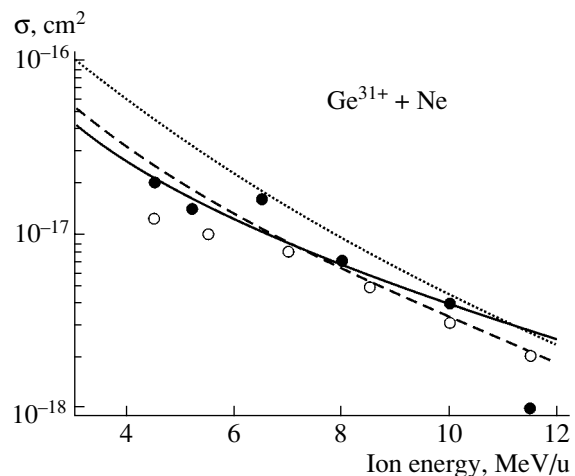

Fig. 6. Total cross sections of charge exchange between H-like Ge ions and Ne atoms, $Ge^{31+} + Ne \longrightarrow Ge^{30+}(n) + Ne^+$, vs. Ge^{31+} ion energy: ●, data points [17]; ○, calculation by the Monte Carlo method [16]; dotted line, eikonal approximation [17]; dashed line, Schlachter semiempirical formula [18]; and continuous curve, CAPTURE code (this study).
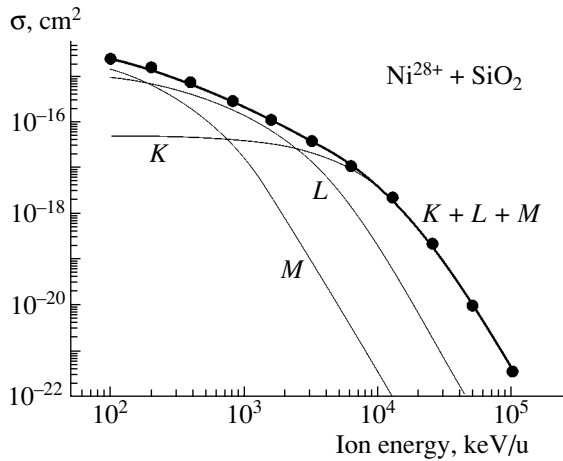
Fig. 7. Cross sections of charge exchange between bare Ni nuclei and SiO_2 molecules, $Ni^{28+} + SiO_2 \longrightarrow Ni^{27+} +$ $[SiO_2]^+$, vs. ion energy for the zero target density ($\rho_t \longrightarrow 0$). K, L, and M are the contributions from the corresponding SiO_2 atomic shells. $K + L + M$ is the total charge exchange cross section (calculation in this study).

Fig. 8. The same as in Fig. 7 for the target atom density $\rho_t = 7 \times 10^{22}$ cm^{-3} (calculation in this study).

quartz, 2.23 g/cm^3 ($\rho_t = 6.7 \times 10^{22}$ cm^{-3}). The purpose of the investigation is to study the effect of density on the energy losses of incident ions.

In this study, we calculated the charge exchange cross sections by formulas (12), (14), and (15) in view of the effect of the target density. Figure 7 shows the total charge exchange cross sections for reaction (19) in the case of a low-density target ($\rho_t \longrightarrow 0$) that were calculated with the cutoff parameter $n_{cut} = \infty$ and with regard to electron capture from all SiO_2 shells. At energies $E = 100$ keV/u–1 MeV/u, electrons are seen to be captured largely from the L and M shells of SiO_2 to the levels $n = 1$–5 of hydrogen-like Ni^{27+} ions. At $E > 10$ MeV/u, only K electrons of the target participate in the process.

As the target becomes denser, the pattern changes radically. For $\rho_t = 7.0 \times 10^{22}$ cm^{-3} (Fig. 8), charge exchange proceeds through the capture of only K elec-
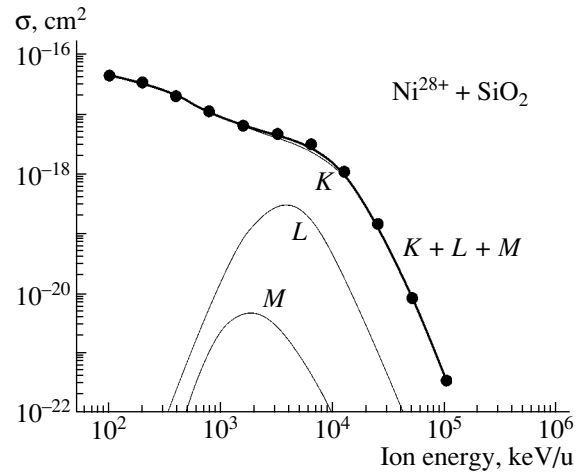
trons in the entire energy range and the total cross section decreases by more than one order of magnitude.

Figure 9 compares the total cross sections of charge exchange between Ni^{28+} and SiO_2 that were calculated for a small ($\rho_t \longrightarrow 0$, curve *1*) and high ($\rho_t = 7 \times 10^{22}$ cm^{-3}) SiO_2 molecule density. With the energy E varying in the range 100–1000 keV/u, taking into account the density effect leads to a decrease in the charge exchange cross section by more than one order of magnitude. At higher energies, $E > 10$ MeV/u, the cross sections are nearly equal to each other, since in both cases only K electrons of the target are captured mostly to the levels $n \approx 1$ and 2 in a Ni^{27+} ion.

In the case of complex targets, the distribution of resulting ions over n, in general, strongly depends on the electronic configuration of the target: at high collision energies, the capture of inner-shell electrons dominates over charge exchange associated with outer shells. The capture of inner electrons greatly complicates calculations even if the target density is low. It may so happen that the cross section distribution over n become nonmonotonic. This is shown in Fig. 10, where

Table 2. Relative cross sections $\sigma_n / \sum_{n=3}^{n=5} \sigma_n$ for the charge exchange reaction $U^{90+} + A \longrightarrow U^{89+}(n) + A^+$ ($A = N_2$, Ar, Kr, and Xe) at the collision energy $E = 162$ MeV/u

n	N_2			Ar			Kr			Xe		
	experiment	theory	this study	experiment	theory	this study	experiment	theory	this study	experiment	theory	this study
3	0.634	0.602	0.556	0.644	0.525	0.471	0.619	0.533	0.501	0.620	0.541	0.519
4	0.253	0.263	0.328	0.240	0.299	0.322	0.257	0.296	0.309	0.256	0.292	0.300
5	0.113	0.135	0.116	0.116	0.176	0.207	0.124	0.171	0.190	0.124	0.167	0.181

Note: The sum of the reduced cross sections σ_n is normalized to unity: $\sum_{n=3}^{n=5} \sigma_n = \sigma_3 + \sigma_4 + \sigma_5 = 1$.
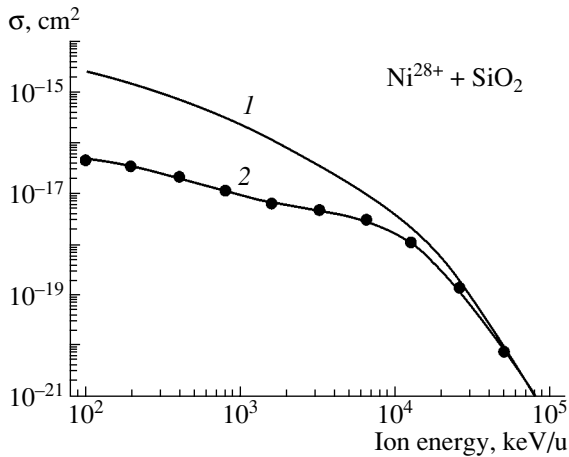
Fig. 9. Cross sections of charge exchange between Ni ions and SiO$_2$ molecules, Ni^{28+} + SiO$_2$ ⟶ Ni^{27+} + [SiO$_2$]$^+$, vs. ion energy for (*1*) zero target density and (*2*) target atom density $\rho_t = 5 \times 10^{22}$ cm^{-3}.

Fig. 10. Distribution of the cross sections of the charge exchange reaction Ni^{28+} + SiO$_2$ ⟶ Ni^{27+}(n) + [SiO$_2$]$^+$ vs. the principal quantum number n of hydrogen-like ions Ni^{27+}(n) for the energy $E = 200$ keV/u and zero target density.

the cross sections of charge exchange (19) are distributed over quantum numbers n for $E = 200$ keV/u and zero target density. The capture of K electrons generates lower lying states Ni^{27+}(n) with the principal quantum numbers $n = 1$ and 2, while charge exchange involving the outer L and M shells of SiO$_2$ brings nickel ions into highly excited states with $n > 5$.

As was noted earlier, experimental data for partial cross sections σ_{nl} and σ_{nlj} of charge exchange between multiply charged ions are scarce. Table 2 compares the relative cross sections σ_n of charge exchange bringing ions to the states with $n = 3$, 4, and 5 for the reaction U^{90+} + A ⟶ U^{89+}(n) + A^+ (A = N$_2$, Ar, Kr, and Xe) at the high collision energy $E = 162$ MeV/u ($v = 80.5$ a.u.). The partial cross sections for charge exchange resulting in the states $3lj$, $4lj$, and $5lj$ were experimentally found [10] from the X-ray emission spectra of Li-like uranium ions by the technique described above. All the cross sections have a distinct peak corresponding to the generation of the final ion in the p state, i.e., with the orbital quantum number $l = 1$. The CAPTURE code, which was used in this work, makes it possible to compute only those cross sections that are averaged over the quantum numbers l and j; therefore, Table 2 lists only the cross sections σ_n. The data in Table 2 are seen to be consistent with each other. The cross sections found in this work take into account the capture of electrons from all target shells and the density effect.

It should be noted that the cross sections of charge exchange between positive ions and atoms are usually measured for low densities of target atoms, $\rho_t = 10^{12}$–10^{15} cm^{-3}, when the density effect is weak. Sometimes, however, taking this effect into account provides a better fit to experimental data, which indirectly proves its
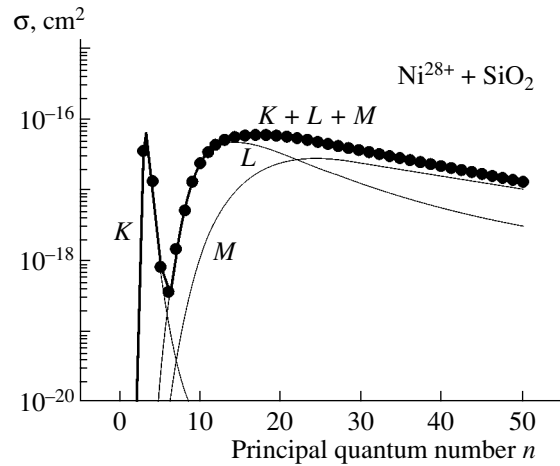
significance. For example, work [5] gives an explanation of the experimentally measured mean (equilibrium) charge of fast uranium ions U^{q+} with a charge $q = 28$–80 and energy $E = (3.6$–11.5) MeV/u when they interact with the dense plasma ($\rho_t \approx 5 \times 10^{19}$ cm^{-3}) of polyethylene (CH)$_n$. It turned out that, with the plasma density included, that is, with the use of the effective (decreased) charge exchange cross sections, the calculated and experimentally found mean charges of uranium ions were in better agreement: $q_{theor} = 60 \pm 2$ versus $q_{expt} = 63 \pm 1$. Thus, measurements of ion–atom charge exchange cross sections for the case of high-density targets would allow one to substantiate the effect of target density on the charge exchange cross sections and spectral emission line intensities of resulting ions.

CONCLUSIONS

The effect of target density on the cross sections of charge exchange between fast ions X^{q^+} and atoms or molecules was studied. The total charge exchange cross sections decrease with increasing target density because target atoms ionize resulting excited ions $[X^{(q-1)^+}]^*$. The indirect effect of target density is confirmed by experimental and calculated data for the mean charge of ion beams passing through a plasma target. Direct measurements of partial and total charge exchange cross sections, as well as the emission line intensities of resulting ions, as a function of the target density allow one to check the validity of this effect.

REFERENCES

1. H. D. Betz, Rev. Mod. Phys. **44**, 465 (1972).

2. H. D. Betz, in *Applied Atomic Collision Physics*, Vol. 4: Condensed Matter (Academic, London, 1983), Part 2.

3. H. Tawara, P. Richard, U. I. Safronova, *et al.*, Phys. Rev. A **64**, 042712 (2001).

4. O. Rosmej, F. Rosmej, J. Wieser, *et al.*, Laser Part. Beams **20**, 479 (2002).

5. H. Wahl, M. Geissel, A. A. Golubev, *et al.*, Kratk. Soobshch. Fiz. 8, 28 (2001).

6. S. Tugarinov, M. Von Hellermann, I. Beigman, *et al.*, in *Proceedings of the Conference on Advanced Diagnostics for Magnetic and Inertial Fusion*, Varenna, Italy, 2001.

7. L. Bureyeva, V. Lisitsa, V. Sergeev, *et al.*, in *Proceedings of the 3rd US–Japan Plasma Polarization Spectroscopy Workshop*, Ed. by P. Beiersdorfer and T. Fujimoto (University of California Press, Livermore, 2001).

8. P. Rymuza, Th. Stöhlker, H. Geissel, *et al.*, Acta Phys. Pol. B **27**, 573 (1996).

9. Th. Stöhlker, F. Bosch, R. W. Dunford, *et al.*, Phys. Scr. T **80**, 379 (1999).

10. X. Ma, Th. Stöhlker, F. Bosch, *et al.*, Phys. Rev. A **64**, 012704 (2001).

11. L. P. Presnyakov, V. P. Shevelko, and R. K. Janev, Elementary Processes with the Participation of Multiply Charged Ions (Atomizdat, Moscow, 1986).

12. R. Mann, F. Folkmann, and H. Beyer, J. Phys. B **14**, 1161 (1981).

13. V. P. Shevel'ko, Zh. Tekh. Fiz. **71** (10), 20 (2001) [Tech. Phys. **46**, 1225 (2001)].

14. V. P. Shevelko, I. Yu. Tolstikhina, and Th. Stöhlker, Nucl. Instrum. Methods B **184**, 295 (2001).

15. I. I. Sobelman, L. A. Vainshtein, and E. A. Yukov, *Excitation of Atoms and Broadening of Spectral Lines*, 2nd ed. (Springer, Berlin, 1995).

16. R. E. Olson, Phys. Rev. A **24**, 1726 (1981).

17. Th. Stöhlker, Ch. Kozhuharov, P. H. Mokler, *et al.*, J. Phys. B **25**, 4527 (1992).

18. A. S. Schlachter, J. W. Stears, W. G. Graham, *et al.*, Phys. Rev. A **27**, 3372 (1983).

Translated by V. Isaakyan

GASES
AND LIQUIDS

On the Thermophoresis of a Spheroidal Solid Aerosol Particle

N. V. Malaĭ and E. R. Shchukin

Belgorod State University, Belgorod, 308015 Russia
e-mail: malay@bsu.edu.ru
Received November 13, 2002; in final form, February 26, 2003

Abstract—The steady motion of a spheroidal aerosol particle with inner nonuniformly distributed heat sources (sinks) that is placed in an external temperature gradient is theoretically studied in the Stokes approximation. The mean temperature of the particle surface is assumed to differ slightly from that of the gaseous environment. An analytic expression for the force and rate of thermophoresis are found by solving the gas-dynamic equations in view of the motion of the environment. © *2003 MAIK "Nauka/Interperiodica".*

1. FORMULATION OF THE PROBLEM

To date, the thermophoresis of spherical aerosol particles has been studied in great detail [1–4]. Many particles occurring both in nature and in industrial plants are nonspherical, e.g., spheroidal. The problem of thermophoresis of a spheroidal aerosol particle has been considered in [5–7]. However, the convective terms in the heat conduction equation were neglected. Praudman and Pearson [8] for the hydrodynamic problem, as well as Acrivos and Taylor [9] for the heat problem, showed that, away from the sphere, the inertial and convective terms became of the same order of magnitude as the molecular-transport terms. Therefore, normal expansion in a small parameter introduces an error, since it fails to rigorously satisfy the boundary conditions at infinity and find a unique exact solution that is valid throughout the flow region even in a second approximation. From the above, it follows that the issue of how the motion of the medium affects the force and rate of thermophoresis of a spheroidal particle is of theoretical and practical interest.

Consider the steady motion of a spheroidal (oblate spheroid) solid aerosol particle with a velocity **U** in the negative direction of the $0z$ axis. The particle contains nonuniformly distributed heat sources with a density q_i. The gas is at rest at infinity, and a small temperature gradient ∇T is provided by external sources. The temperature drop in the neighborhood of the particle is assumed to be small; i.e., $(T_s - T_\infty)/T_\infty \ll 1$, where T_s is the mean temperature of the particle surface and T_∞ is the gas temperature away from the particle. In this case, the thermal conductivity, as well as the dynamic and kinematic viscosity, can be considered as constants and the gas, as an incompressible medium. The particle size is considerably larger than the free paths of gas mixture molecules; therefore, corrections in Knudsen number will be neglected [3]. Hereafter, the subscripts e and i correspond to the environment and spheroid, respectively.

We will describe the thermophoresis of a particle in the spheroidal coordinate system (ε, η, φ) with the origin at the center of the spheroid; i.e., the origin of the fixed coordinate system coincides with the instantaneous position of the center of the particle. The curvilinear coordinates ε, η, and φ are related to the Cartesian coordinates by the relations [10]

$$x = c \sinh\varepsilon \sin\eta \cos\varphi, \quad y = c \sinh\varepsilon \sin\eta \sin\varphi, \tag{1.1}$$
$$z = c \cosh\varepsilon \cos\eta,$$

$$x = c \cosh\varepsilon \sin\eta \cos\varphi, \quad y = c \cosh\varepsilon \sin\eta \sin\varphi, \tag{1.2}$$
$$z = c \sinh\varepsilon \cos\eta,$$

where $c = \sqrt{b^2 - a^2}$ in the case of a prolate spheroid ($a < b$, formula (1.1)) or $c = \sqrt{a^2 - b^2}$ for an oblate spheroid ($a > b$, formula (1.2)) and a and b are the spheroid semiaxes. The $0z$ axis of the Cartesian coordinate system coincides with the symmetry axis of the spheroid.

In view of the above assumptions, the distributions of the velocity $\mathbf{U}_e$, pressure P_e, and temperatures T_e and T_i are described by the set of equations (1.3) and (1.4) with boundary conditions (1.5)–(1.7):

$$\nabla P_e = \mu_e \Delta \mathbf{U}_e, \quad \text{div}\,\mathbf{U}_e = 0, \tag{1.3}$$

$$\rho_e c_{pe}(\mathbf{U}_e \cdot \nabla)T_e = \lambda_e \Delta T_e, \quad \Delta T_i = -q_i/\lambda_i, \tag{1.4}$$

$$U_\varepsilon = -\frac{cU\cosh\varepsilon}{H_\varepsilon}\cos\eta,$$

$$U_\eta = \frac{cU\sinh\varepsilon}{H_\varepsilon}\sin\eta - K_{tc}\frac{\nu_e}{T_e}(\nabla T_e \cdot \mathbf{e}_\eta), \tag{1.5}$$

$$T_e = T_i, \quad \lambda_e(\nabla T_e \cdot \mathbf{e}_\varepsilon) = \lambda_i(\nabla T_i \cdot \mathbf{e}_\varepsilon) \quad \text{for} \quad \varepsilon = \varepsilon_0,$$

$$\mathbf{U}_e \longrightarrow 0, \quad T_e \longrightarrow T_\infty + |\nabla T|c\sinh\varepsilon\cos\eta,$$

$$P_e \longrightarrow P_\infty \quad \text{for} \quad \varepsilon \longrightarrow \infty, \tag{1.6}$$

$$T_i \neq \infty \quad \text{for} \quad \varepsilon \longrightarrow 0. \tag{1.7}$$

Here, $\mathbf{e}_\varepsilon$ and $\mathbf{e}_\eta$ are the unit vectors of the spheroidal coordinate system; λ is the thermal conductivity; $U = |\mathbf{U}|$; $H_\varepsilon = \sqrt{\cosh^2\varepsilon - \sin^2\eta}$ is the Lamé coefficient; c_{pe} is the specific heat; K_{tc} is the thermal creep coefficient, which is calculated from the kinetic theory of gases; and ∇ is the Laplacian. Today, the most rigorous expression for the coefficient K_{tc} is known for a spherical particle [3]. The gas-kinetic coefficient $K_{tc} = 1.152$ when the accommodation coefficients of tangential momentum, α_τ, and energy, α_E, are equal to unity [3, 4]. In numerical calculations, we assume that the coefficient K_{tc} for a spheroid differs insignificantly from that for a sphere [6].

Boundary conditions (1.5) on the particle surface ($\varepsilon = \varepsilon_0$) allow for creep for the tangent component of the mass velocity, temperature equality, and the continuity of heat fluxes on the particle surface. Away from the particle ($\varepsilon \longrightarrow \infty$), boundary conditions (1.6) are valid, and the finiteness of the physical quantities characterizing the particle at $\varepsilon \longrightarrow 0$ is taken into account in (1.7).

The resultant force acting on a spheroidal particle from the environment is given by the formula [11]

$$F_z = \int_S \left(-P_e\cos\eta + \sigma_{\varepsilon\varepsilon}\cos\eta - \frac{\sinh\varepsilon}{\cosh\varepsilon}\sigma_{\varepsilon\eta}\sin\eta \right) dS, \tag{1.8}$$

where $dS = c^2\cosh^2\varepsilon\,\sin\eta\,d\eta\,d\varphi$ is a differential element of area, and $\sigma_{\varepsilon\varepsilon}$ and $\sigma_{\varepsilon\eta}$ are the strain tensor components in the spheroidal coordinate system.

2. TEMPERATURE DISTRIBUTION IN THE VICINITY OF THE PARTICLE, FORCE AND VELOCITY OF THERMOPHORESIS

We make Eqs. (1.3) and (1.4) and boundary conditions (1.5)–(1.7) dimensionless by introducing the dimensionless velocity, temperature, and pressure: $\mathbf{V}_e = \mathbf{U}_e/U$, $t_k = T_k/T_\infty$, and $p_k = P_k/P_\infty$ $(k = e, i)$. Here, the spheroid major semiaxis is taken as the unit length; U, as the unit velocity; $P_\infty = \mu_e U/a$, as the unit pressure; and T_∞, as the unit temperature $(U \sim \mu_e|\nabla T|/(\rho_e T_\infty))$.

Expressions (1.3)–(1.7) have the single controllable small parameter $\xi = a|\nabla T|/T_\infty \ll 1$. Therefore, we will look for a solution to the boundary-value problem in the form of expansion in powers of ξ:

$$\mathbf{V}_e = \mathbf{V}_e^{(0)} + \xi\mathbf{V}_e^{(1)} + \dots, \quad t = t^{(0)} + \xi t^{(1)} + \dots,$$
$$p_e = p_e^{(0)} + \xi p_e^{(1)} + \dots. \tag{2.1}$$

We will restrict our consideration to the first-order terms in ξ when calculating the force acting on the particle and the velocity of its thermophoretic motion in the given external temperature gradient field. In order

to find these quantities, one has to know the distributions of the velocity, pressure, and temperature both outside and inside the spheroid. Substituting (2.1) into (1.4), leaving terms $\sim\xi$, and solving the sets of equations found by the method of separation of variables, we will finally find in the zero approximation $(\xi = 0)$

$$t_e^{(0)}(\lambda) = 1 + \gamma\lambda_0\operatorname{arccot}\lambda \quad (\lambda = \sinh\varepsilon), \tag{2.2}$$

$$t_i^{(0)}(\lambda) = D + \frac{\lambda_e}{\lambda_i}\gamma\lambda_0\operatorname{arccot}\lambda$$
$$+ \int_{\lambda_0}^{\lambda}\operatorname{arccot}\lambda f\,d\lambda - \operatorname{arccot}\lambda\int_{\lambda_0}^{\lambda}f\,d\lambda. \tag{2.3}$$

Here, $\lambda_0 = \sinh\varepsilon_0$, $\gamma = t_s - 1$ is the dimensionless parameter, $t_s = T_s/T_\infty$, and T_s is the mean temperature of the spheroid surface given by

$$\frac{T_s}{T_\infty} = 1 + \frac{1}{4\pi\lambda_e c\lambda_0 T_\infty}\int_V q_i\,dV,$$

$$D = 1 + \left(1 - \frac{\lambda_e}{\lambda_i}\right)\gamma\lambda_0\operatorname{arccot}\lambda_0, \tag{2.4}$$

$$f = -\frac{c^2}{2\lambda_i T_\infty}\int_{-1}^{+1}q_i(\lambda^2 + x^2)dx, \quad x = \cos\eta.$$

In (2.4), the integral is taken over the entire particle volume. In the first approximation $(\sim\xi)$,

$$t_e^{(1)}(\lambda, x) = \cos\eta\left\{\frac{c\lambda}{a} + \Gamma c(\lambda\operatorname{arccot}\lambda - 1)\right.$$
$$+ \omega\left[A_2\left(\operatorname{arccot}\lambda - \frac{\lambda}{2}\operatorname{arccot}^2\lambda\right)\right. \tag{2.5}$$
$$\left.\left. + \frac{A_1}{2}(\operatorname{arccot}\lambda - \lambda\operatorname{arccot}^2\lambda)\right]\right\},$$

$$t_i^{(1)}(\lambda, x) = \cos\eta\left\{Bc\lambda + \frac{3(1 - \lambda\operatorname{arccot}\lambda)}{4\pi c^2\lambda_i T_\infty}\int_V q_i z\,dV\right.$$
$$\left. - \lambda\int_{\lambda_0}^{\lambda}(\lambda\operatorname{arccot}\lambda - 1)f_1\,d\lambda + (\lambda\operatorname{arccot}\lambda - 1)\int_{\lambda_0}^{\lambda}\lambda f_1\,d\lambda\right\}. \tag{2.6}$$

Here, $\omega = \Pr\gamma\lambda_0/(ac)$ and $\Pr$ is the Prandtl number. The constants of integration A_1 and A_2 appear in expressions for the components of the mass velocity and pressure. These expressions are found by solving Stokes equations (1.3) in the oblate coordinate system and have the

form [10]

$$U_\varepsilon(\varepsilon, \eta) = \frac{U}{c\cosh\varepsilon H_\varepsilon}\cos\eta \qquad (2.7)$$
$$\times \{\lambda A_2 + [\lambda - (1 + \lambda^2)\mathrm{arccot}\lambda]A_1 + c^2(1 + \lambda^2)\},$$

$$U_\eta(\varepsilon\eta) = -\frac{U}{cH_\varepsilon}\sin\eta\left\{\frac{A_2}{2} + (1 - \lambda\,\mathrm{arccot}\lambda)A_1 + c^2\lambda\right\}, \qquad (2.8)$$

$$P_e(\varepsilon, \eta) = P_\infty + c\frac{\mu_e U}{H_\varepsilon^4}x(\lambda^2 + x^2)A_2. \qquad (2.9)$$

The constants Γ and B enter into expressions (2.5) and (2.6) for the temperature fields inside and outside the particle from the corresponding boundary conditions on the spheroid surface. Since an expression for the coefficient Γ will be of interest to us, we write it in explicit form:

$$\Gamma = -\frac{1 - \delta}{\Delta a} + \frac{3}{4\pi c^3 \lambda_i T_\infty \Delta\lambda_0(1 + \lambda_0^2)}\int_V q_i z dV$$

$$+ \frac{\omega}{c\Delta}\left\{A_2\left[-\frac{\delta}{1 + \lambda_0^2} + \left(\delta\frac{\lambda_0^2}{1 + \lambda_0^2} - \frac{1}{\lambda_0}\right)\mathrm{arccot}\lambda_0\right.\right.$$

$$+ \frac{1 - \delta}{2}\mathrm{arccot}^2\lambda_0\bigg] + \frac{A_1}{2}\bigg[-\frac{\delta}{1 + \lambda_0^2} + (1 - \delta)\,\mathrm{arccot}^2\lambda_0 \qquad (2.10)$$

$$\left.\left. + \left(\delta\frac{2\lambda_0}{1 + \lambda_0^2} - \frac{1}{\lambda_0}\right)\mathrm{arccot}\lambda_0\right]\right\},$$

$$\Delta = (1 - \delta)\mathrm{arccot}\lambda_0 + \delta\frac{\lambda_0}{1 + \lambda_0^2} - \frac{1}{\lambda_0}, \quad \delta = \frac{\lambda_e^s}{\lambda_i^s}.$$

Hereafter, the superscript s denotes the values of physical quantities at the mean temperature T_s of the spheroid surface, which is given by formula (2.4).

Substituting (2.7)–(2.9) into (1.8) and integrating, we arrive at

$$F_z = 4\pi\frac{\mu_e U}{c}A_2. \qquad (2.11)$$

The coefficient A_2 is found from boundary conditions (1.5) in view of expressions (2.7), (2.8), and (2.10):

$$A_2 = -\frac{2c^2}{\beta[\lambda_0 + (1 - \lambda_0^2)\mathrm{arccot}\lambda_0]}$$

$$- 2K_{tc}\frac{v_e^s c^2}{t_s U}\frac{\delta}{1 + \lambda_0^2}\frac{\lambda_0 - (1 + \lambda_0^2)\mathrm{arccot}\lambda_0}{\beta[\lambda_0 + (1 - \lambda_0^2)\mathrm{arccot}\lambda_0]}\frac{|\nabla T|}{T_\infty}$$

$$\times\left\{1 - \frac{3a(\lambda_0\,\mathrm{arccot}\lambda_0 - 1)}{4\pi c^3 \lambda_e T_\infty\lambda_0}\int_V q_i z dV\right.$$

$$\left. + \mathrm{Pr}\frac{1 + \lambda_0^2}{2}\lambda_0\gamma\frac{1 - \lambda_0\,\mathrm{arccot}\lambda_0(2 - \lambda_0\,\mathrm{arccot}\lambda_0)}{\lambda_0 - (1 + \lambda_0^2)\mathrm{arccot}\lambda_0}\right\},$$

where

$$\beta = 1 - 2K_{tc}\frac{\lambda_0}{t_s}\gamma\frac{\mathrm{Pr}}{\Delta}\frac{\delta}{1 + \lambda_0^2}\frac{\lambda_0 - (1 + \lambda_0^2)\mathrm{arccot}\lambda_0}{\lambda_0 + (1 - \lambda_0^2)\mathrm{arccot}\lambda_0}$$

$$\times\left[1 - \left(\lambda_0 + \frac{1}{2}\mathrm{arccot}\lambda_0\right)\mathrm{arccot}\lambda_0\right.$$

$$\left. - \frac{\lambda_0 - (2 - \lambda_0\,\mathrm{arccot}\lambda_0)\lambda_0^2\,\mathrm{arccot}\lambda_0}{\lambda_0 - (1 + \lambda_0^2)\mathrm{arccot}\lambda_0}\right].$$

In view of the explicit form of the coefficient A_2, we find a general expression for the force acting on a spheroidal particle. This force is the sum of the viscous force F_μ and the force $F^{(1)}$

$$F = F_\mu + F^{(1)}, \qquad (2.12)$$

where

$$F_\mu = -8\pi\mu_e^s U\frac{c^2}{\beta[\lambda_0 + (1 - \lambda_0^2)\mathrm{arccot}\lambda_0]}, \qquad (2.13)$$

$$F^{(1)} = -8\pi\mu_e^s c K_{tc}\frac{v_e^s}{t_s}\frac{\lambda_0 - (1 + \lambda_0^2)\mathrm{arccot}\lambda_0}{\beta[\lambda_0(1 - \lambda_0^2)\mathrm{arccot}\lambda_0]}\frac{\delta}{(1 + \lambda_0^2)\Delta}$$

$$\times\frac{|\nabla T|}{T_\infty}\left[1 - \frac{3a(\lambda_0\mathrm{arccot}\lambda_0 - 1)}{4\pi c^3 \lambda_e T_\infty\lambda_0}\int_V q_i z dV\right. \qquad (2.14)$$

$$\left. + \mathrm{Pr}\frac{1 + \lambda_0^2}{2}\lambda_0\gamma\frac{1 - \lambda_0\,\mathrm{arccot}\lambda_0(2 - \lambda_0\,\mathrm{arccot}\lambda_0)}{\lambda_0 - (1 + \lambda_0^2)\mathrm{arccot}\lambda_0}\right].$$

In the general case, the force $F^{(1)}$ is the sum of three forces: the thermophoretic force, the force proportional to the dipole moment of the density of heat sources nonuniformly distributed over the particle volume, and the third term due to the motion of the medium (i.e., the force component taking into account the convective terms in the heat conduction equation).

Equating the resultant force F to zero, we arrive at a general expression for the drift (thermophoretic) velocity of a solid oblate spheroidal particle in the external temperature gradient field:

$$U_{th} = -\frac{b}{a}K_{tc}\frac{v_e^s}{t_s}\delta$$

$$\times \frac{1 - (\lambda_0 + 1/\lambda_0)\operatorname{arccot}\lambda_0}{\sqrt{1 + \lambda_0^2}\left[(1 - \delta)\operatorname{arccot}\lambda_0 + \delta\dfrac{\lambda_0}{1 + \lambda_0^2} - \dfrac{1}{\lambda_0}\right]}$$

$$(2.15)$$

$$\times \left[1 - \frac{3a}{4\pi c^3 \lambda_e T_\infty}\left(\operatorname{arccot}\lambda_0 - \frac{1}{\lambda_0}\right)\int_V q_i z dV\right.$$

$$\left. + \Pr\frac{1 + \lambda_0^2}{2}\lambda_0\gamma\frac{1 - \lambda_0\operatorname{arccot}\lambda_0(2 - \lambda_0\operatorname{arccot}\lambda_0)}{\lambda_0 - (1 + \lambda_0^2)\operatorname{arccot}\lambda_0}\right]\frac{|\nabla T|}{T_\infty}.$$

In order to find the rate of thermophoresis for a prolate spheroid, one has to substitute $i\lambda$ for λ and ic for c (i is the imaginary unit) in (2.15).

Thus, formulas (2.12) and (2.15) have the most general form and make it possible to estimate the resultant force acting on a solid spheroidal aerosol particle and its drift velocity in the external temperature gradient field for the case when heat sources (sinks) are nonuniformly distributed inside the particle. In this approach, the motion of the environment is taken into account for small temperature differences in the vicinity of the particle.

3. RESULTS AND DISCUSSION

If one does not take into account the motion of the environment and internal heat sources, (2.15) is reduced to an expression for the purely thermophoretic velocity of a spheroidal particle:

$$U_{th}^b = K_{tc}\nu_e\delta f_{th}^{(b)}\frac{|\nabla T|}{T_\infty}$$

$$\left(f_{th}^{(b)} = -\frac{b}{a}\frac{1 - (\lambda_0 + 1/\lambda_0)\operatorname{arccot}\lambda_0}{\sqrt{1 + \lambda_0^2}\left[(1 - \delta)\operatorname{arccot}\lambda_0 + \delta\dfrac{\lambda_0}{1 + \lambda_0^2} - \dfrac{1}{\lambda_0}\right]}\right),$$

$$(3.1)$$

which coincides with formula (9) in [5].

In the case of a sphere, (2.15) turns into an expression for the thermophoretic velocity of a solid spherical particle of radius R that includes the flow of the environment and internal heat sources:

$$U^{(sph)}(a = b = R) = K_{tc}\nu_e^s\delta f^{(sph)}\frac{|\nabla T|}{T_\infty},\qquad (3.2)$$

where

$$\gamma_0 = \frac{1}{4\pi R\lambda_e T_\infty}\int_V q_i dV,$$

$$f^{(sph)} = -\frac{2}{t_e^s(1 + 2\delta)}\left\{1 + \frac{1}{4\pi R^2 \lambda_e T_\infty}\int_V q_i dV - \frac{\Pr}{12}\gamma_0\right\}.$$

Disregarding the flow of the environment and internal heat sources yields the conventional formula for the thermophoretic velocity of a large spherical particle [1, 2]

$$U_{th}(a = b = R) = -2K_{tc}\frac{\nu_e^s}{t_e^s}\frac{\delta}{1 + 2\delta}\frac{|\nabla T|}{T_\infty}.\qquad (3.3)$$

In order to estimate how the motion of the environment affects the thermophoretic velocity of a spheroidal particle, one has to specify the nature of heat sources nonuniformly distributed over its volume. As an example, let us consider the simplest case when the particle absorbs radiation as a black body. In this case, radiation is absorbed in a thin layer of depth $\delta\varepsilon \ll \varepsilon_0$ that is adjacent to the heated particle surface. The density of heat sources inside the layer of depth $\delta\varepsilon$ is equal to [12, 13]

$$q_i(\varepsilon, \eta) = \begin{cases} -\dfrac{\cosh\varepsilon\cos\eta}{c(\cosh^2\varepsilon - \sin^2\eta)\delta\varepsilon}I_0, & \dfrac{\pi}{2} \le \eta \le \pi, \\ \varepsilon_0 - \delta\varepsilon \le \varepsilon \le \varepsilon_0 \\ 0, & 0 \le \eta \le \dfrac{\pi}{2}, \end{cases}\qquad (3.4)$$

where I_0 is the intensity of an incident radiation.

The integrals $\int_V q_i dV$ and $\int_V q_i z dV$ appear in the expression for the thermophoretic velocity. Substituting (3.4) into these integrals in view of the fact that $\delta\varepsilon \ll \varepsilon_0$ and performing integration, we find

$$\int_V q_i dV = \pi I_0 c^2 \lambda_0^2\left(1 + \frac{1}{\lambda_0^2}\right),$$

$$(3.5)$$

$$\int_V q_i z dV = -\frac{2}{3}\pi I_0 c^3 \lambda_0^3\left(1 + \frac{1}{\lambda_0^2}\right).$$

In view of (3.5), expression (2.15) takes the form

$$U_{th}^* = K_{tc}\nu_e^s\delta f_{th}^*\frac{|\nabla T|}{T_\infty},\qquad (3.6)$$

where

$$f_{th}^* = -\frac{b}{a}\frac{1 - (\lambda_0 + 1/\lambda_0)\operatorname{arccot}\lambda_0}{\sqrt{1 + \lambda_0^2}t_e^s\left[(1 - \delta)\operatorname{arccot}\lambda_0 + \delta\dfrac{\lambda_0}{1 + \lambda_0^2} - \dfrac{1}{\lambda_0}\right]}$$

$$\times \left\{1 + \frac{\lambda_0^2 a}{2\lambda_e T_\infty}I_0\left(1 + \frac{1}{\lambda_0^2}\right)[\lambda_0\operatorname{arccot}\lambda_0 - 1\right.\qquad (3.7)$$

$$\left. + \frac{\Pr}{4}\sqrt{1 + \lambda_0^2}\frac{1 - \lambda_0\operatorname{arccot}\lambda_0(2 - \lambda_0\operatorname{arccot}\lambda_0)}{\lambda_0 - (1 + \lambda_0^2)\operatorname{arccot}\lambda_0}\right]\right\}.$$

Table 1

a, μm	$b/a = 0.1$							
	$I_0 \times 10^2$, W/m^2							
	0.5		2		5		10	
	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$
15	0.52	0.32	2.08	1.30	5.20	3.29	10.40	6.71
20	0.69	0.43	2.77	1.74	6.93	4.42	13.87	9.08
25	0.87	0.54	3.47	2.18	8.67	5.56	17.33	11.51

Table 2

a, μm	$b/a = 0.3$							
	$I_0 \times 10^2$, W/m^2							
	0.5		2		5		10	
	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$
15	0.39	0.20	1.59	0.81	3.97	2.04	7.95	4.16
20	0.59	0.27	2.12	1.08	5.30	2.74	10.60	5.63
25	0.66	0.33	2.65	1.35	6.62	3.44	13.25	7.13

Table 3

a, μm	$b/a = 0.5$							
	$I_0 \times 10^2$, W/m^2							
	0.5		2		5		10	
	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$
15	0.32	0.12	1.26	0.48	3.15	1.20	6.31	2.45
20	0.42	0.16	1.68	0.64	4.21	1.61	8.41	3.32
25	0.53	0.20	2.10	0.80	5.26	2.03	10.51	4.20

In the case of a sphere, (3.6) is recast as

$$U_{th}^{(sph)} = K_{tc} v_e^s \delta f_{th}^{(sph)} \frac{|\nabla T|}{T_\infty}, \tag{3.8}$$

where

$$f^{(sph)} = -\frac{2}{t_e^s(1 + 2\delta)}\left[1 - \frac{RI_0}{6\lambda_e T_\infty}\left(1 + \frac{Pr}{8}\right)\right]. \tag{3.9}$$

The mean temperature of the spheroid surface is related to the incident radiation intensity I_0 as

$$T_s = T_\infty + \frac{c\lambda_0}{4\lambda_e}I_0\left(1 + \frac{1}{\lambda_0^2}\right). \tag{3.10}$$

In order to illustrate the contributions of the form-factor (ratio of the spheroid semiaxes), flow of the environment, and internal heat release (nonuniform distribution of heat sources over the particle volume) to the thermophoretic velocity (3.6), Tables 1–4 list the numerical estimations for particles of borated graphite ($\lambda_i^s = 55$ W/(m K)) suspended in air at $T_\infty = 280$ K and $P_e = 10^5$ Pa.

The numerical analysis showed that, at a given ratio between the semiaxes, the relative contribution of the other factors increases with increasing incident radiation intensity I_0. This effect depends significantly on the equatorial radius of the spheroid (a). For instance, in Table 1 ($a = 15$ μm), $f^{(1)} = 0.52$ at $I_0 = 0.5 \times 10^2$ W/m^2 and $f^{(1)} = 10.40$ at $I_0 = 10 \times 10^2$ W/m^2 ($f^{(1)} = (|f_{th}^* - f_{th}^{(b)}|/f_{th}^{(b)}) \times 100\%$). Such behavior of the function $f^{(1)}$ is due to the fact that, as follows from (3.10), (3.7), and the numerical estimations, the major contribution is from the terms proportional to the dipole moment of the density of heat sources nonuniformly distributed over the particle volume. In (3.7), this is the term

Table 4

a, μm	$b/a = 0.8$							
	$I_0 \times 10^2$, W/m^2							
	0.5		2		5		10	
	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$	$f^{(1)}$	$f^{(2)}$
15	0.24	0.04	0.94	0.15	2.35	0.38	4.69	0.77
20	0.31	0.05	1.25	0.20	3.13	0.51	6.26	1.04
25	0.39	0.06	1.56	0.25	3.91	0.64	7.82	1.32

$\lambda_0 \operatorname{arccot} \lambda_0 - 1$. The dimensionless term related to the motion of the environment (see the heat conduction equation) is proportional to the Prandtl number. In a gas, this number is on the order of unity; therefore, the contribution of this term differs from that of the first one by one order of magnitude. This fact may be used to separate particles by size, finely purify gases from aerosol particles, estimate translucent zones appearing in clouds and fogs when they are probed by laser radiation, etc. The influence of the factors mentioned above will increase with increasing radiation intensity. However, the mean temperature of the spheroid surface will also increase (see (3.10)). In this case, we cannot consider the coefficients of molecular transport to be constant. Therefore, expressions (2.15) and (3.6) must involve the mean values of the physical quantities at a given temperature of the particle surface, which is determined by (2.4) and (3.10), to avoid large errors. It is also of interest to compare the thermophoretic velocity with that for a spherical particle with a radius equal to the equatorial radius of a spheroid, i.e., with formula (3.8). The numerical analysis showed that in this case, too, the relative error increases with increasing incident radiation intensity and equatorial radius. For example, in Table 1 ($a = 15$ μm), $f^{(2)} = 0.32$ at $I_0 = 0.5 \times 10^2$ W/m^2 and $f^{(2)} = 6.71$ at $I_0 = 10 \times 10^2$ W/m^2 ($f^{(2)} = (|f_{th}^{sph} - f_{th}^{(b)}|/f_{th}^{(b)}) \times 100\%$). However, this increase is approximately 1.5 times smaller than that in the former case.

REFERENCES

1. S. P. Bakanov and B. V. Deryagin, Dokl. Akad. Nauk SSSR **142**, 139 (1962).
2. Yu. I. Yalamov and A. S. Sanasaryan, Inzh.-Fiz. Zh. **28**, 1061 (1975).
3. Yu. I. Yalamov, A. B. Poddoskin, and A. A. Yushkanov, Dokl. Akad. Nauk SSSR **254**, 1047 (1980) [Sov. Phys. Dokl. **25**, 734 (1980)].
4. A. B. Poddoskin, A. A. Yushkanov, and Yu. I. Yalamov, Zh. Tekh. Fiz. **52**, 2253 (1982) [Sov. Phys. Tech. Phys. **27**, 1383 (1982)].
5. K. H. Leong, J. Aerosol Sci. **15**, 511 (1984).
6. Yu. I. Yalamov, V. P. Redchits, and M. N. Gaĭdukov, Inzh.-Fiz. Zh. **39**, 538 (1980).
7. A. A. Gukasyan, Thermophoresis of a Moderately Coarse Spherical Pareticle, Available from VINITI, 1981, Moscow, No. 5320-81.
8. I. Praudman and J. R. A. Pearson, J. Fluid Mech. **2**, 237 (1957).
9. A. Acrivos and T. D. Taylor, J. Phys. Fluids **5**, 387 (1962).
10. J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics* (Prentice-Hall, Englewood Cliffs, 1965; Mir, Moscow, 1976).
11. L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Gostekhizdat, Moscow, 1958; Oxford, Pergamon, 1959).
12. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983; Mir, Moscow, 1986).
13. S. Bretshnaĭder, *Properties of Gases and Liquids: Engineering Methods of Analysis* (Moscow, Khimiya, 1966).

Translated by M. Fofanov

GASES
AND LIQUIDS

Nonlinear Capillary Vibrations of a Charged Drop Placed in a Dielectric Medium: Single-Mode Initial Deformation of the Drop Shape

V. A. Koromyslov, S. O. Shiryaeva, and A. I. Grigor'ev

Demidov State University, Sovetskaya ul. 14, Yaroslavl, 150000 Russia
e-mail: shir@uniyar.ac.ru
Received February 7, 2003

Abstract—The nonlinear vibrations of the equilibrium spherical shape of a charged drop placed in a perfect incompressible dielectric medium are asymptotically calculated in the second-order approximation in single-mode initial deformation of the drop surface. The drop is assumed to be a perfect incompressible liquid. It is shown that the nonlinear vibration amplitudes, as well as the energy distribution between nonlinearly excited modes, depend significantly on the parameter ρ, where ρ is the ratio of the environmental density to that of the drop. It is also demonstrated that an increase in ρ raises the amplitude of the highest of the vibration modes excited due to second-order nonlinear interaction. In the second order of smallness, the amplitude of the zeroth mode is independent of the density ratio. As ρ grows, the effect of the self-charge of the drop, the interfacial tension, and the permittivity of the environment on the nonlinear oscillations increases. © 2003 MAIK "Nauka/Interperiodica".

(1) The study of electrostatic instability of a charged liquid drop suspended in another liquid is of considerable interest for numerous applications, such as uniform mixing of immiscible liquids, combustion of liquid fuels to mix a fuel with an oxidant, and geophysical experiments (see, for instance, [1–4] and references cited there). Nevertheless, many related issues are as yet little understood because of the experimental and theoretical complexity of the problem. In the majority of cases, it has been solved in the linear approximation in amplitude of capillary vibrations. A number of recent theoretical and numerical works [5–11] devoted to the calculation of charged drop vibration in a vacuum take into account the actual nonlinearity of the phenomenon. In [12], the nonlinear vibrations of a charged conducting drop placed in a compressible dielectric medium were studied only for the initial disturbance of the fundamental vibration mode, and the emphasis was on the acoustic radiation of the drop. Since in most applications the drop vibration takes place in media that cannot be identified as vacuum [1–4], it seems to be reasonable to study the influence of the environment on the nonlinear vibrations of the drop in a more general situation where the initial deformation of the equilibrium spherical shape is defined by the virtual excitation of an arbitrary single mode of capillary vibrations.

(2) Consider a system consisting of two immiscible ideal incompressible liquids with densities ρ_1 and ρ_2. In the absence of the gravitational field, the inner liquid (related parameters will be marked by subscript 1) takes the shape of a spherical drop with a radius R under the action of interfacial tension forces with a coefficient σ.

The outer liquid (subscript 2) is assumed to be unbounded. Let the drop liquid be ideally conducting and bear a charge Q distributed over the surface and the environment be an ideal dielectric with a permitivity ε_*. We also assume that, at zero time $t = 0$, the equilibrium spherical shape of the drop undergoes a virtual axisymmetric perturbation of a given amplitude which is much smaller than the drop radius and is proportional to the amplitude of one of the capillary vibration modes. Our aim is to find an analytical expression for the shape of the nonlinearly vibrating drop as a function of time ($t > 0$).

Below, we use dimensionless variables such that $R = \rho_1 = \sigma = 1$ and assume that $\rho_2/\rho_1 \equiv \rho$.

In the spherical coordinate system (r, Θ, ϕ) with the origin at the drop center, the equation of the interface disturbed by axisymmetric capillary wave motion has the form

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

The motion of the liquid in the drop and environment is potential; that is, the velocity fields of the wave motion in the drop, $\mathbf{V}(\mathbf{r}, t) = \nabla\psi(\mathbf{r}, t)$, and in the surrounding medium, $\mathbf{U}(\mathbf{r}, t) = \nabla\varphi(\mathbf{r}, t)$, are entirely defined by the velocity potential functions $\psi(\mathbf{r}, t)$ and $\varphi(\mathbf{r}, t)$.

In the above statement, a set of equations for interface evolution consists of the Laplace equations for the velocity potentials $\psi(\mathbf{r}, t)$ and $\varphi(\mathbf{r}, t)$ and electrostatic potential $\Phi(\mathbf{r}, t)$:

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

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

$$\Delta\Phi(\mathbf{r}, t) = 0 \qquad (4)$$

(here, ∇ is the Laplacian).

The boundary conditions are as follows:

at the center of the drop,

$$r \longrightarrow 0: \psi(\mathbf{r}, t) \longrightarrow 0; \qquad (5)$$

at infinity,

$$r \longrightarrow \infty: \Phi(\mathbf{r}, t) \longrightarrow 0; \quad \varphi(\mathbf{r}, t) \longrightarrow 0; \qquad (6)$$

at the interface:

$$r = 1 + \xi(\Theta, t); \quad \frac{\partial \xi}{\partial t} = \frac{\partial \psi}{\partial r} - \frac{1}{r^2}\frac{\partial \xi}{\partial \Theta}\frac{\partial \psi}{\partial \Theta} \qquad (7)$$

(kinematic condition),

$$\frac{\partial \varphi}{\partial r} - \frac{1}{r^2}\frac{\partial \xi}{\partial \Theta}\frac{\partial \varphi}{\partial \Theta} = \frac{\partial \psi}{\partial r} - \frac{1}{r^2}\frac{\partial \xi}{\partial \Theta}\frac{\partial \psi}{\partial \Theta} \qquad (8)$$

(equality of the velocity normal components of the liquid in the drop and in the medium),

$$-\frac{\partial \psi}{\partial t} - \frac{1}{2}(\nabla\psi)^2 + P_0^{(1)} + P_E - P_\sigma$$
$$= -\rho\frac{\partial \varphi}{\partial t} - \frac{\rho}{2}(\nabla\varphi)^2 + P_0^{(2)} \qquad (9)$$

(dynamic condition),

$$\Phi(\mathbf{r}, t) = \Phi_S(t) \qquad (10)$$

(constancy of the electric potential at the drop surface).

In the mathematical relationships above, $P_0^{(j)}$ is the pressure in the drop ($j = 1$) and in the medium ($j = 2$), $P_E = \varepsilon_*(\nabla\Phi)^2/8\pi$ is the electric field pressure upon the interface, $P_\sigma = \mathrm{div}\,\mathbf{n}$ is the Laplacian pressure ($\mathbf{n}$ is the unit vector of the positive normal to the drop surface, and $\Phi_S(t)$ is the constant electrostatic potential on the drop surface.

One must also take into account the constancy of the electric charge,

$$-\frac{1}{4\pi}\oint_S \varepsilon_*(\mathbf{n}\cdot\nabla\Phi)dS = Q, \quad S = \begin{cases} r = 1 + \xi(\Theta, t) \\ 0 \le \Theta \le \pi \\ 0 \le \phi \le 2\pi; \end{cases} \qquad (11)$$

constancy of the drop volume,

$$\int_{V_1} r^2 dr \sin\Theta d\Theta d\phi = \frac{4}{3}\pi; \quad V_1 = \begin{cases} 0 \le r \le 1 + \xi(\Theta, t) \\ 0 \le \Theta \le \pi \\ 0 \le \phi \le 2\pi; \end{cases} \qquad (12)$$

and immovability of the center-of-mass of the drop,

$$\frac{\rho_1\int_{V_1}\mathbf{r}dV_1 + \rho_2\int_{V_2}\mathbf{r}dV_2}{\int_{V_1}\rho_1 dV_1 + \int_{V_2}\rho_2 dV_2} = 0; \quad V_2 = \begin{cases} r > 1 + \xi(\Theta, t) \\ 0 \le \Theta \le \pi \\ 0 \le \phi \le 2\pi. \end{cases} \qquad (13)$$

The initial conditions to the problem stated are the initial axisymmetric deformation of the equilibrium spherical shape of the drop,

$$t = 0: \xi(\Theta) = \xi_0 + \varepsilon P_k(\mu) \; (k \ge 2); \; \mu \equiv \cos\Theta \qquad (14)$$

and the zero initial interface velocity,

$$t = 0: \frac{\partial \xi(\Theta, t)}{\partial t} = 0. \qquad (15)$$

In (14) and (15), ε is a small parameter that has the meaning of the initial deformation amplitude; $P_k(\mu)$ is the Legendre polynomial of degree k; and ξ_0 is a constant determined from condition (12) accurate to the second order of smallness,

$$\xi_0 = -\varepsilon^2\frac{1}{(2k+1)} + O(\varepsilon^3). \qquad (16)$$

(3) To find a solution to the problem in an approximation quadratic in ε, we use the well-known method of many scales [13]. For this purpose, we represent the desired functions $\xi(\Theta, t)$, $\psi(\mathbf{r}, t)$, $\varphi(\mathbf{r}, t)$, and $\Phi(\mathbf{r}, t)$ in the form of series in powers of the small parameter ε and assume that these functions depend not merely on time t but on various time scales T_m that are defined via the small parameter $T_m \equiv \varepsilon^m t$:

$$\xi(\Theta, t) = \sum_{m=0} \varepsilon^m \xi^{(m)}(\Theta, T_0, T_1, \ldots);$$

$$\psi(\mathbf{r}, t) = \sum_{m=0} \varepsilon^m \psi^{(m)}(\Theta, T_0, T_1, \ldots);$$

$$\varphi(\mathbf{r}, t) = \sum_{m=0} \varepsilon^m \varphi^{(m)}(\Theta, T_0, T_1, \ldots);$$

$$\Phi(\mathbf{r}, t) = \sum_{m=0} \varepsilon^m \Phi^{(m)}(\Theta, T_0, T_1, \ldots). \qquad (17)$$

The derivatives with respect to time will be taken over the whole set of time scales by the rule

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial T_0} + \varepsilon\frac{\partial}{\partial T_1} + O(\varepsilon^2).$$

Substituting expansions (17) into boundary-value problem (2)–(13) and equating terms of the same order of smallness in each of the equations, one easily obtains a set of boundary-value problems for the successive

determination of the unknown functions $\xi^{(m)}$, $\psi^{(m)}$, $\varphi^{(m)}$, and $\Phi^{(m)}$, where $m = 0, 1, 2,\ldots$.

Due to the linearity of Eqs. (2)–(6), each of the functions $\psi^{(m)}$, $\varphi^{(m)}$, and $\Phi^{(m)}$ in expansions (17) must satisfy them; therefore, we represent these functions as solutions to set (2)–(6) in the form

$$\psi^{(m)}(r, \Theta, T_0, T_1, \ldots) = \sum_{n=0}^{\infty} D_n^{(m)}(T_0, T_1, \ldots)r^n P_n(\mu);$$

$$\varphi^{(m)}(r, \Theta, T_0, T_1, \ldots) = \sum_{n=0}^{\infty} G_n^{(m)}(T_0, T_1, \ldots)r^{-(n+1)}P_n(\mu); \quad (18)$$

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

The successive corrections $\xi^{(m)}$ to the expression for the drop shape will be sought in the form of Legendre polynomials:

$$\xi^{(m)}(r, \Theta, T_0, T_1, \ldots) = \sum_{n=0}^{\infty} M_n^{(m)}(T_0, T_1, \ldots)P_n(\mu). \quad (19)$$

(4) In the zeroth order of smallness in ε, we obtain a solution for the equilibrium state of the system:

$$\varphi^{(0)}(r, \Theta, T_0, T_1, \ldots) \equiv 0; \quad \psi^{(0)}(r, \Theta, T_0, T_1, \ldots) \equiv 0;$$
$$\xi^{(0)}(r, \Theta, T_0, T_1, \ldots) \equiv 0; \quad \Phi^{(0)}(r, \Theta, T_0, T_1, \ldots) = \frac{Q}{r\varepsilon_*}. \quad (20)$$

(5) To find the coefficients $G_n^{(1)}$, $D_n^{(1)}$, $F_n^{(1)}$, and $M_n^{(1)}$ in solutions (18) and (19) (at $m = 1$) in the first-order approximation in ε, we transform boundary conditions (7)–(13) into the form

$$r = 1: \frac{\partial \xi^{(1)}}{\partial T_0} = \frac{\partial \psi^{(1)}}{\partial r};$$

$$\frac{\partial \varphi^{(1)}}{\partial r} = \frac{\partial \psi^{(1)}}{\partial r};$$

$$-\frac{\partial \psi^{(1)}}{\partial T_0} + \frac{\varepsilon_*}{4\pi}\frac{d\Phi^{(0)}}{dr}\left(\frac{\partial \Phi^{(1)}}{\partial r} + \frac{d^2\Phi^{(0)}}{\partial r^2}\xi^{(1)}\right)$$

$$+ (2 + \Delta_\Omega)\xi^{(1)} = -\rho\frac{\partial \varphi^{(1)}}{\partial T_0};$$

(21)

$$\Phi^{(1)} + \frac{d\Phi^0}{dr}\xi^{(1)} = \Phi_S^{(1)};$$

$$\int_0^\pi \left[\frac{\partial \Phi^{(1)}}{\partial r} + \left(\frac{d^2\Phi^{(0)}}{dr^2} + 2\frac{d\Phi^{(0)}}{dr}\right)\xi^{(1)}\right]d\mu = 0;$$

$$\int_0^\pi \xi^{(1)}d\mu = 0; \quad \int_\Omega \xi^{(1)}Y_1^n(\Theta, \varphi)d\Omega = 0;$$

$$\Delta_\Omega \equiv \frac{\partial}{\partial\mu}\left((1 - \mu^2)\frac{\partial}{\partial\mu}\right),$$

where $Y_1^n(\Theta, \varphi)$ is a spherical function.

Expressions for the first-order coefficients in expansions (18) and (19) are easily found from set (21):

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

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

$$D_n^{(1)}(T_0, T_1, \ldots) = \frac{1}{n}\frac{\partial M_n^{(1)}(T_0, T_1, \ldots)}{\partial T_0} \quad (n \geq 2);$$

$$G_n^{(1)}(T_0, T_1, \ldots) = -\frac{n}{n+1}D_n^{(1)}(T_0, T_1, \ldots)$$

$$= -\frac{1}{n+1}\frac{\partial M_n^{(1)}(T_0, T_1, \ldots)}{\partial T_0} \quad (n \geq 2); \quad (22)$$

$$F_n^{(1)}(T_0, T_1, \ldots) = \frac{Q}{\varepsilon_*}M_n^{(1)}(T_0, T_1, \ldots) \quad (n \geq 0);$$

$$A_n(T_1, \ldots) \equiv a_n(T_1, \ldots)\exp[ib_n(T_1, \ldots)];$$

$$\Phi_S^{(1)} = 0; \quad \omega_n^2 \equiv \kappa_n n(n-1)[(n+2) - W];$$

$$W \equiv \frac{Q^2}{4\pi\varepsilon_*}; \quad \kappa_n = \left(1 + \frac{n\rho}{n+1}\right)^{-1}.$$

The dependences of the functions a_n and b_n on the parameters T_1, etc. are determined in the next orders of smallness.

To complete the consideration of the problem in the linear approximation in ε, we assume that the quantities a_n and b_n are independent of time; that is,

$$a_n(T_1, \ldots) \approx a_n^{(0)} + O(T_1);$$

$$b_n(T_1, \ldots) \approx b_n^{(0)} + O(T_1).$$

Then, it is easy to check that the shape perturbation can be estimated as

$$\xi(\Theta, t) \approx \varepsilon\xi^{(1)}(\Theta, t, a_n^{(0)}, b_n^{(0)}) + \varepsilon O(\varepsilon t), \quad (23)$$

where $a_n^{(0)}$ and $b_n^{(0)}$ are time-invariable constants.

The error in expansion (23) turns out to be on the order of the first term if $t \approx O(\varepsilon^{-1})$. For $t \geq O(\varepsilon^{-1})$, expansion (23) becomes invalid. Thus, expression (23) is valid over the time interval $t \leq O(1)$, and, in this case, the error is on the order of $\sim\varepsilon^2$. However, when studying the motion of the surface, one may also use (23) at

$t \le O(\varepsilon^{-1})$ provided that the first-order solution is comparable to the initial perturbation amplitude. More detailed information on the applicability of this expansion may be obtained from the second-order analysis of the problem.

Substituting expansion (23) into initial conditions (14)–(16) subject to (19) and (22) and equating the quantities of the same order of smallness to each other yields

$$a_n^{(0)} = \frac{1}{2}\delta_{nk}; \quad b_n^{(0)} = 0 \ (n \ge 2),$$

where δ_{nk} is the Kronecker symbol.

Eventually, the function describing the drop shape evolution with time in the linear (in ε) approximation has the form

$$\xi(\Theta, t) = \varepsilon\cos(\omega_k t)P_k(\mu) + O(\varepsilon^2). \quad (24)$$

According to this first-order solution, the drop surface harmonically vibrates about the equilibrium sphere at the kth (initially excited) mode.

For the velocity field potentials and electrostatic potential, we obtain, in the same approximation, the following solutions:

$$\psi(\mathbf{r}, t) = -\varepsilon\frac{\omega_k r^k}{k}\sin(\omega_k t)P_k(\mu) + O(\varepsilon^2);$$

$$\varphi(\mathbf{r}, t) = \varepsilon\frac{\omega_k}{(k+1)r^{k+1}}\sin(\omega_k t)P_k(\mu) + O(\varepsilon^2); \quad (25)$$

$$\Phi(\mathbf{r}, t) = \frac{Q}{r\varepsilon_*} + \varepsilon\cos(\omega_k t)\frac{Q}{r^{k+1}\varepsilon_*}P_k(\mu) + O(\varepsilon^2).$$

(6) To find second-order corrections to the above solutions (that is, to find the functions $\xi^{(2)}(\Theta, t)$, $\psi^{(2)}(\mathbf{r}, t)$, $\varphi^{(2)}(\mathbf{r}, t)$, and $\Phi^{(2)}(\mathbf{r}, t)$, we write a set of equations that follows from boundary conditions (7)–(13) by equating terms $\sim\varepsilon^2$:

$$r = 1: \frac{\partial\xi^{(2)}}{\partial T_0} + \frac{\partial\xi^{(1)}}{\partial T_1} = \frac{\partial\psi^{(2)}}{\partial r} + \frac{\partial^2\psi^{(1)}}{\partial r^2}\xi^{(1)} - \frac{\partial\xi^{(1)}}{\partial\Theta}\frac{\partial\psi^{(1)}}{\partial\Theta};$$

$$\frac{\partial\psi^{(2)}}{\partial r} + \frac{\partial^2\psi^{(1)}}{\partial r^2}\xi^{(1)} - \frac{\partial\xi^{(1)}}{\partial\Theta}\frac{\partial\psi^{(1)}}{\partial\Theta}$$

$$= \frac{\partial\varphi^{(2)}}{\partial r} + \frac{\partial^2\varphi^{(1)}}{\partial r^2}\xi^{(1)} - \frac{\partial\xi^{(1)}}{\partial\Theta}\frac{\partial\varphi^{(1)}}{\partial\Theta};$$

$$-\frac{\partial\psi^{(2)}}{\partial T_0} - \frac{\partial\psi^{(1)}}{\partial T_1} - \frac{\partial^2\psi^{(1)}}{\partial r\partial T_0}\xi^{(1)} - \frac{1}{2}\left[\left(\frac{\partial\psi^{(1)}}{\partial r}\right)^2 + \left(\frac{\partial\psi^{(1)}}{\partial\Theta}\right)^2\right]$$

$$+ \frac{\varepsilon_*}{8\pi}\left[2\frac{d\Phi^{(0)}}{dr}\frac{\partial\Phi^{(2)}}{\partial r} + \left(\frac{\partial\Phi^{(1)}}{\partial r}\right)^2 + \left(\frac{\partial\Phi^{(1)}}{\partial\Theta}\right)^2\right.$$

$$+ 2\frac{d\Phi^{(0)}}{dr}\frac{d^2\Phi^{(0)}}{dr^2}\xi^{(2)} + 2\left(\frac{d\Phi^{(0)}}{dr}\frac{\partial^2\Phi^{(1)}}{\partial r^2} + \frac{d^2\Phi^{(0)}}{dr^2}\frac{\partial\Phi^{(1)}}{\partial r}\right)\xi^{(1)}$$

$$+ \left(\frac{d\Phi^{(0)}}{dr}\frac{d^3\Phi^{(0)}}{dr^3} + \left(\frac{d^2\Phi^{(0)}}{dr^2}\right)^2\right)(\xi^{(1)})^2\right] + (2 + \Delta_\Omega)\xi^{(2)}$$

$$(26)$$

$$- 2\xi^{(1)}(1 + \Delta_\Omega)\xi^{(1)} = -\rho\left\{\frac{\partial\varphi^{(2)}}{\partial T_0} + \frac{\partial\varphi^{(1)}}{\partial T_1}\right.$$

$$+ \frac{\partial^2\varphi^{(1)}}{\partial r\partial T_0}\xi^{(1)} + \frac{1}{2}\left[\left(\frac{\partial\varphi^{(1)}}{\partial r}\right)^2 + \left(\frac{\partial\varphi^{(1)}}{\partial\Theta}\right)^2\right]\right\};$$

$$\Phi^{(2)} + \frac{d\Phi^{(0)}}{dr}\xi^{(2)} + \frac{\partial\Phi^{(1)}}{\partial r}\xi^{(1)} + \frac{1}{2}\frac{d^2\Phi^{(0)}}{dr^2}(\xi^{(1)})^2 = \Phi_S^{(2)};$$

$$\int_0^\pi\left[\frac{d\Phi^{(2)}}{dr} + \left(\frac{\partial^2\Phi^{(1)}}{\partial r^2} + 2\frac{\partial\Phi^{(1)}}{\partial r}\right)\xi^{(1)} + \left(\frac{d^2\Phi^{(0)}}{dr^2} + 2\frac{d\Phi^{(0)}}{dr}\right)\xi^{(2)}\right.$$

$$+ \left(\frac{1}{2}\frac{d^3\Phi^{(0)}}{dr^3} + 2\frac{d^2\Phi^{(0)}}{dr^2} + \frac{d\Phi^{(0)}}{dr}\right)(\xi^{(1)})^2 - \frac{\partial\Phi^{(1)}}{\partial\Theta}\frac{\partial\xi^{(1)}}{\partial\Theta}\right]d\mu = 0;$$

$$\int_0^\pi[\xi^{(2)}(\Theta, t) + (\xi^{(1)}(\Theta, t))^2]d\mu = 0;$$

$$\int_0^{2\pi}\int_0^\pi[2\xi^{(2)}(\Theta, t) + 3(\xi^{(1)}(\Theta, t))^2]Y_1^n(\Theta, \varphi)d\mu d\varphi = 0.$$

Substituting expansions (18) and (19) (at $m = 2$), as well as solutions (20) and (22), into set (26) of the boundary conditions yields differential equations for the unknown coefficients $M_n^{(2)}(T_0, T_1, \ldots)$. From the requirement that the secular terms be eliminated from the solutions, we obtain

$$\frac{\partial a_n}{\partial T_1} = 0, \quad \frac{\partial b_n}{\partial T_1} = 0;$$

this means that a_n and b_n in solutions (21) are independent of the time scale T_1. Their dependence on the slower time scales T_2, T_3, etc., can be determined only in higher order approximations.

The general solution to inhomogeneous differential equations obtained from set (26) is written in the form

$$M_0^{(2)}(T_0, T_2, \ldots) = -\sum_{n=2}^\infty\frac{1}{(2n+1)}\{A_n(T_2, \ldots)\bar{A}_n(T_2, \ldots)$$

$$+ [A_n(T_2, \ldots)]^2\exp(i2\omega_n T_0) + \text{c.c.}\};$$

$$M_1^{(2)}(T_0, T_2, \ldots) = -9 \sum_{n=2}^{\infty} \frac{n}{(2n+1)(2n-1)}$$
$$\times \{ A_{n-1}(T_2, \ldots) A_n(T_2, \ldots) \exp(i\omega_{n-1} + \omega_n) T_0 \quad (27)$$
$$+ \overline{A}_{n-1}(T_2, \ldots) A_n(T_2, \ldots) \exp(i(\omega_n - \omega_{n-1}) T_0) + \text{c.c.} \};$$

$$M_n^{(2)}(T_0, T_1, \ldots) = \{ c_n(T_2, \ldots) \exp(id_n(T_1, \ldots)) $$
$$\times \exp(i\omega_n T_0) + \text{c.c.} \} + N_n(T_0, T_2, \ldots).$$

Also we have

$$N_n(T_0, T_2, \ldots) \equiv \sum_{n=2}^{\infty} \sum_{l=2}^{\infty} \{ \lambda_{mln}^{(+)} A_m(T_2, \ldots) A_l(T_2, \ldots)$$
$$\times \exp[i(\omega_m + \omega_l) T_0] + \lambda_{mln}^{(-)} A_m(T_2, \ldots)$$
$$\times \overline{A}_l(T_2, \ldots) \exp[i(\omega_m - \omega_l) T_0] + \text{c.c.} \};$$

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

$$\gamma_{mln} \equiv \kappa_n K_{mln} \left\{ \omega_m^2 \left[n - m + 1 - \frac{\rho n(n-m-1)}{n+1} \right] \right.$$
$$+ 2n[l(l+1) - 1] + W\frac{n}{2}[l(m+1) - m(2m-2n+7) + 3] \right\}$$
$$+ \kappa_n \alpha_{mln} \left\{ \omega_m^2 \left[\frac{1}{m} - \frac{n\rho}{(n+1)(m+1)} \right] + W\frac{n}{2} \right\};$$

$$\eta_{mln} \equiv \kappa_n K_{mln} \left\{ \frac{n}{2} - m + 1 + \frac{\rho n(2m-n+3)}{2(n+1)} \right\}$$
$$+ \kappa_n \alpha_{mln} \left\{ \frac{1}{m}\left(1 + \frac{n}{2l}\right) - \frac{n\rho(n+2l+3)}{2(m+1)(l+1)(n+1)} \right\};$$

$$K_{mln} \equiv [C_{mln}^{000}]^2;$$

$$\alpha_{mln} \equiv -\sqrt{m(m+1)l(l+1)} C_{mln}^{000} C_{mln}^{-110};$$

$$C_{mln}^{000} \equiv \begin{cases} 0, & \text{if} \quad m+l+n = 2g+1, \\ & \text{where } g \text{ is an integer;} \\ \dfrac{(-1)^{g-n}\sqrt{2n+1}\,g!}{(g-m)!(g-l)!(g-n)!} \\ \times \left[\dfrac{(2g-2m)!(2g-2l)!(2g-2n)!}{(2g+1)!} \right]^{1/2}, \\ \text{if} \quad m+l+n = 2g \;\; (g \text{ is an integer});\end{cases}$$

$$C_{mln}^{-110} \equiv \sqrt{2n+1}\,n!$$
$$\times \left[\frac{(m+l-n)!m(m+1)}{(n+m-l)!(n-m+l)!(m+l+n+1)!l(l+1)} \right]^{1/2}$$
$$\times \sum_z \frac{(-1)^{m+1+z}(m+z-1)!(n+l-z+1)}{z!(m-z+1)!(n-z)!(l-n+z-1)!}.$$

In the last expression, summation is performed over all integer numbers z for which the expressions under the factorial sign are nonnegative. The bar over A_n in (27) denotes the complex conjugate.

Note that C_{mln}^{000} and C_{mln}^{-110} (Clebsch–Gordan coefficients [14]) are other than zero if their subscripts satisfy the relation

$$|m-l| \le n \le (m+l); \quad m+l+n = 2g, \quad (28)$$

where g is an integer.

The coefficients in expansion (19) for the velocity field potentials in the environment, $\varphi^{(m)}(\mathbf{r}, t)$, and drop, $\psi^{(m)}(\mathbf{r}, t)$, and for the electrostatic potentials $\Phi^{(m)}(\mathbf{r}, t)$ are related to solutions (19) and (27) by the relationships

$$D_n^{(2)}(T_0, T_1, \ldots) = \frac{1}{n} \left\{ \frac{\partial M_n^{(2)}(T_0, T_1, \ldots)}{\partial T_0} \right.$$
$$- \sum_{m=2}^{\infty} \sum_{l=2}^{\infty} [(m(m-1)K_{mln} - \alpha_{mln})]$$
$$\left. \times \frac{1}{m} \frac{\partial M_m^{(1)}(T_0, T_2, \ldots)}{\partial T_0} M_l^{(1)}(T_0, T_2, \ldots) \right\};$$

$$G_n^{(2)}(T_0, T_1, \ldots) = \frac{-1}{n+1} \left\{ \frac{\partial M_n^{(2)}(T_0, T_1, \ldots)}{\partial T_0} \right.$$
$$+ \sum_{m=2}^{\infty} \sum_{l=2}^{\infty} [(m+1)(m+2)K_{mln} - \alpha_{mln}]$$
$$\left. \times \frac{1}{m+1} \frac{\partial M_m^{(1)}(T_0, T_2, \ldots)}{\partial T_0} M_l^{(1)}(T_0, T_2, \ldots) \right\};$$

$$F_0^{(2)}(T_0, T_1, \ldots) = 0;$$

$$F_n^{(2)}(T_0, T_1, \ldots) = \frac{Q}{\varepsilon_*} \left\{ M_n^{(2)}(T_0, T_1, \ldots) \right.$$
$$+ \sum_{m=1}^{\infty} \sum_{l=1}^{\infty} m K_{mln} M_m^{(1)}(T_0, T_2, \ldots) M_l^{(1)}(T_0, T_2, \ldots) \right\}.$$

In solution (27), $c_n(T_1, \ldots)$ and $d_n(T_1, \ldots)$ are unknown functions of time that are independent of the time scale T_0.

As in the case of the linear approximation, we complete the quadratic consideration assuming that the unknown values a_n, b_n, c_n, and d_n are independent of time; that is, they are constants defined by initial conditions (14)–(16), because the estimates

$$a_n \approx a_n^{(0)} + O(T_2); \quad b_n \approx b_n^{(0)} + O(T_2);$$

$$c_n \approx c_n^{(0)} + O(T_1); \quad d_n \approx d_n^{(0)} + O(T_1)$$

are valid.

In this case, expansion (19) of the function describing the distortion of the drop shape takes the form

$$\xi(\Theta, t) \approx \varepsilon \xi^{(1)}(\Theta, T_0, a_n^{(0)}, b_n^{(0)})$$

$$+ \varepsilon^2 \xi^{(2)}(\Theta, T_0, a_n^{(0)}, b_n^{(0)}, c_n^{(0)}, d_n^{(0)}) + O(\varepsilon^3 t), \tag{29}$$

where $a_n^{(0)}$, $b_n^{(0)}$, $c_n^{(0)}$, and $d_n^{(0)}$ are constants.

Expression (29) holds over the time interval $t \leq O(1)$ with an error of $\sim\varepsilon^3$. Within the time interval $O(1) < t \leq O(\varepsilon^{-1})$, the error is comparable to the second term (second-order correction); consequently, in expansion (29), only the first (linear) term should be left. Thus, the approximate linear solution of problem (24)–(25) is uniformly applicable over the time interval $t \leq O(\varepsilon^{-1})$.

Substituting expansion (29) into initial conditions (14)–(16) and equating quantities of the same order of smallness to each other yields, in view of (19), the following relationships:

$$t = 0: M_n^{(1)} = \delta_{nk}; \quad M_n^{(2)} = -\frac{\delta_{n0}}{2k+1};$$

$$\frac{\partial M_n^{(1)}}{\partial T_0} = 0; \quad \frac{\partial M_n^{(2)}}{\partial T_0} + \frac{\partial M_n^{(1)}}{\partial T_0} = 0 \ (n = 0, 1, 2, \ldots).$$

Using these relationships and expressions (22) and (27), we find the unknown constants $a_n^{(0)}$, $b_n^{(0)}$, $c_n^{(0)}$, and $d_n^{(0)}$:

$$a_n^{(0)} = \frac{1}{2}\delta_{nk}; \quad b_n^{(0)} = 0;$$

$$c_n^{(0)} = -\frac{(1 - \delta_{n,0})}{2} N_n(t = 0);$$

$$d_n^{(0)} = 0 \ (n = 0, 1, 2, \ldots).$$

Eventually, solutions (22) and (27) for the coefficients $M_n^{(m)}(t)$ take the form

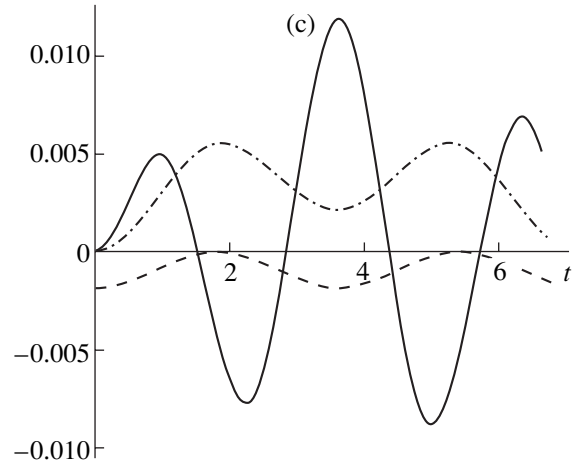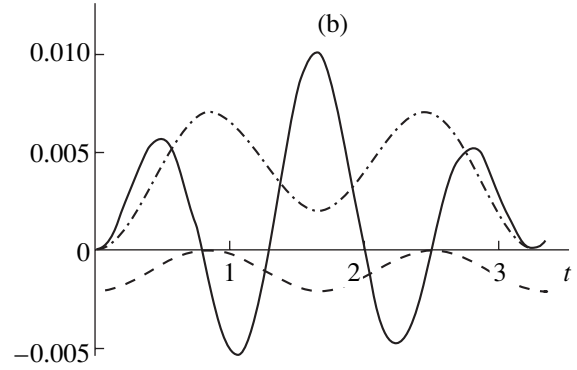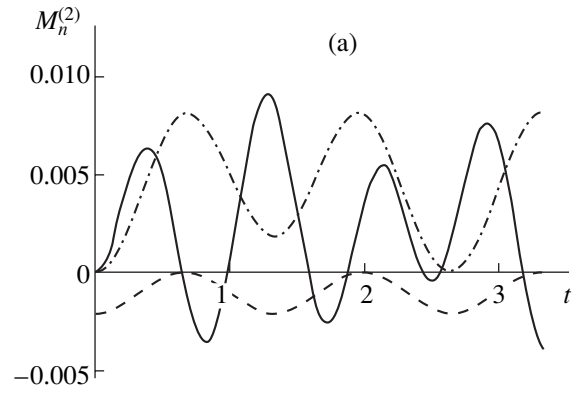$$M_n^{(1)}(t) = \delta_{nk}\cos(\omega_n t);$$

Fig. 1. Time variation of the second-order dimensionless amplitudes $M_n^{(2)}(t)$ when the initial drop deformation is induced by the fundamental mode. $\rho =$ (a) 0.1, (b) 1.0, and (c) 10. $W = 1$.

$$M_0^{(2)}(t) = -\frac{1}{2(2k+1)}[1 + \cos(2\omega_k t)]; \quad M_1^{(2)}(t) = 0; \tag{30}$$

$$M_n^{(2)}(t) = -N_n(0)\cos(\omega_n t) + N_n(t) \ (n \geq 2);$$

$$N_n(t) = \frac{1}{2}(\lambda_{kkn}^{(-)} + \lambda_{kkn}^{(+)}\cos(2\omega_k t)).$$

It is easy to check that the coefficients $M_n^{(2)}(t)$ are proportional to the parameters $\lambda_{kkn}^{(\pm)}$ (which, in turn, are
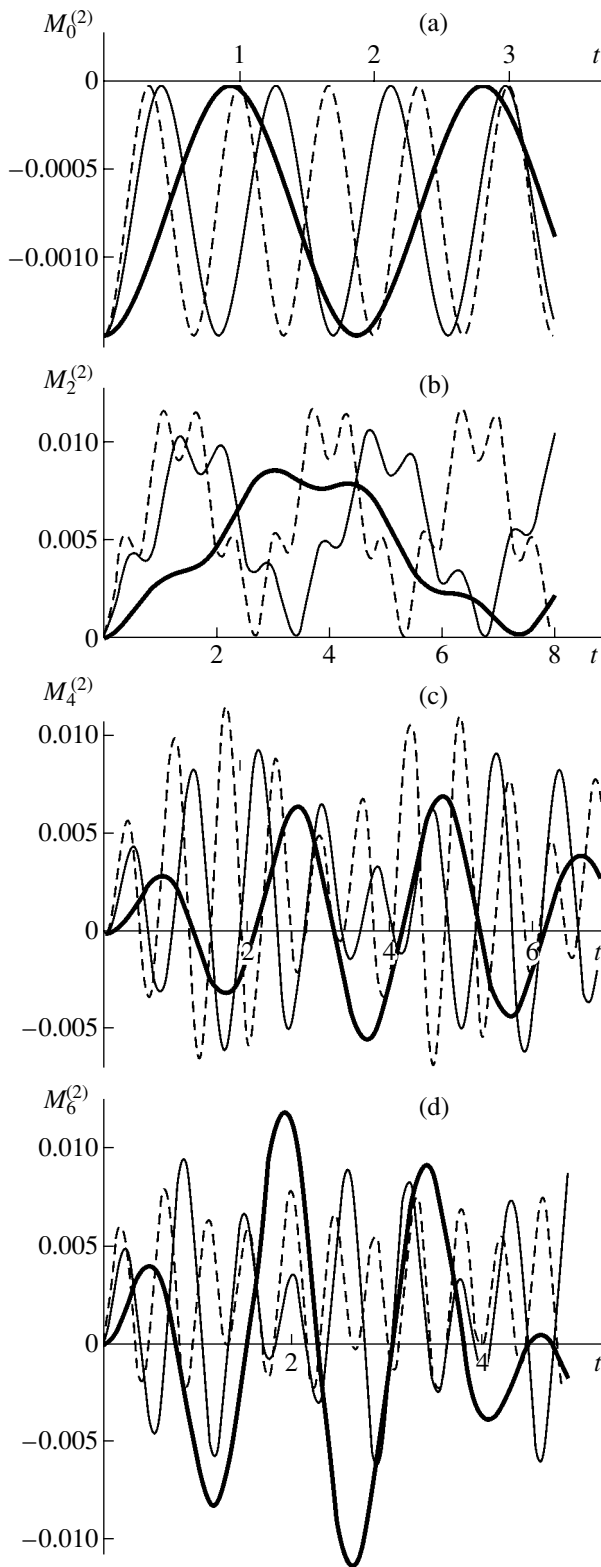
Fig. 2. Time variation of the second-order dimensionless amplitudes of modes $M_n^{(2)}(t)$ when the initial drop deformation is induced by the third mode at $W = 1$. $\rho = 0.1$ (dashed line), 1 (thin solid line), and 10 (thick solid line). (a) The zeroth mode, $n = 0$; (b) the second mode, $n = 2$; (c) the fourth mode, $n = 4$; and (d) the sixth mode, $n = 6$.
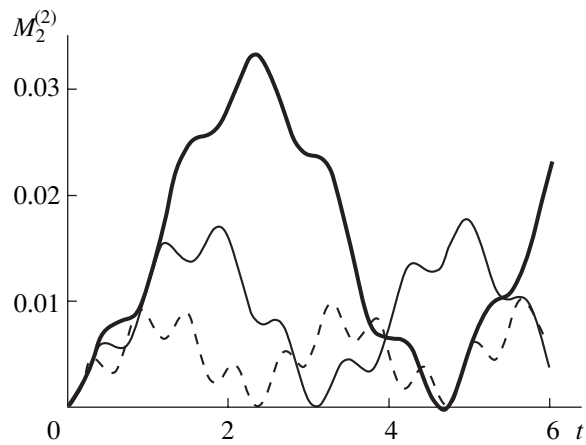

Fig. 3. Time variation of the second-order dimensionless amplitude $M_2^{(2)}(t)$ of the fundamental mode when the initial drop deformation is induced by the third mode at $\rho = 0.1$. $W = 1$ (dashed line), 2 (thin solid line), and 3 (thick solid line).

proportional to the Clebsch-Gordan coefficients C_{kkn}^{000} and C_{kkn}^{-110} and, hence (see (28)), are other than zero only if $n = 2j$ $(j = 0, \ldots, k)$.

Substituting (19) and (30) into (29) yields the second-order approximation for the time variation of the drop shape:

$$
\begin{aligned}
\xi(\Theta, t) &\approx \varepsilon \cos(\omega_k t) P_k(\mu) \\
&- \varepsilon^2 \frac{1}{2} \Bigg\{ \frac{1}{(2k+1)} [1 + \cos(2\omega_k t)] \\
&+ \sum_{j=1}^{k} [(\lambda_{k,k,2j}^{(-)} + \lambda_{k,k,2j}^{(+)}) \cos(\omega_{2j} t) \\
&- (\lambda_{k,k,2j}^{(-)} + \lambda_{k,k,2j}^{(+)} \cos(2\omega_k t))] P_{2j}(\mu) \Bigg\} + O(\varepsilon^3 t).
\end{aligned}
\tag{31}
$$

From (31), it is seen that the initial disturbance of any kth (even or odd) mode of capillary vibrations leads to the excitation (in the second order of smallness) only of even modes whose numbers range between 0 and $2k$.

(7) Figures 1a–1c show the time dependences of the second-order amplitudes of modes excited by the non-linear interaction for the case when the initial deformation is induced by the virtual excitation of the fundamental mode ($k = 2$). The dependences are calculated with (30) at various density ratios ρ. The second-order amplitudes are observed for the zeroth (dashed line), second (dash-and-dot line), and fourth (solid line) modes in this case. The amplitude of the fourth mode grows with increasing density ratio ρ, while that of the second mode declines. When higher modes are responsible for the initial drop deformation, an increase in ρ is always accompanied by the growth of the highest of the modes excited in the second order of smallness, whereas the amplitudes of the remaining modes, except

for the zeroth one, decrease. This fact is illustrated in Figs. 2a–2d, where the calculated results are presented for the case when the initial deformation is associated with the third mode. Each of the panels in Fig. 2 displays the second-order amplitudes of one of the excited modes as a function of time at various density ratios. As was noted above, the amplitude of the zeroth mode does not vary with ρ. This is because the second-order amplitude of the zeroth mode depends on the square of the amplitude of the mode responsible for the initial deformation (see (30)). The ρ dependence of the zeroth mode amplitude appears only in calculations of the third order of smallness.

Figure 3 shows the second-order amplitude of the fundamental mode at $\rho = 0.1$ and various subcritical values of the parameter W, which integrates all relevant physical parameters: the interfacial tension, the permitivity of the environment, and the charge and radius of the drop. In dimensional form, $W = Q^2/4\pi\sigma R^3\varepsilon_*$. It is easy to see that the fundamental mode amplitude grows noticeably as W approaches the critical value $W_{cr} = 4$ [1–4] (above this value, the drop becomes unstable against its self-charge).

From the figures it follows that, as ρ grows, the vibration frequency drops linearly in the small parameter ε. This effect has already been discussed for perfect liquids [15], for viscous liquids and dielectric media [16], and for conductive media [17].

CONCLUSIONS

When a drop of an ideal incompressible conducting liquid immersed in an ideal dielectric incompressible medium nonlinearly vibrates, the energy maximum in the spectrum of nonlinearly excited modes shifts toward the highest mode with growing medium-to-drop density ratio no matter which of the modes is responsible for the initial drop deformation. In the approximation quadratic in the small parameter, the amplitude of the fundamental mode grows when the self-charge of the drop increases and the interfacial tension, drop radius, and permitivity of the medium decrease.

REFERENCES

1. A. G. Baily, Sci. Prog. **61**, 555 (1974).
2. V. I. Kozhenkov and N. A. Fuks, Usp. Khim. **45**, 2274 (1976).
3. A. I. Grigor'ev, Yu. V. Syshchikov, and S. O. Shiryaeva, Zh. Prikl. Khim. (Leningrad) **62**, 2020 (1989).
4. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).
5. J. A. Tsamopoulos and R. A. Brown, J. Fluid Mech. **127**, 519 (1983).
6. O. A. Basaran and L. E. Scriven, Phys. Fluids A **1**, 795 (1989).
7. Z. Feng, J. Fluid Mech. **333**, 1 (1997).
8. S. O. Shiryaeva, Zh. Tekh. Fiz. **71** (2), 27 (2001) [Tech. Phys. **46**, 158 (2001)].
9. S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 163 (2001).
10. S. O. Shiryaeva, Zh. Tekh. Fiz. **72** (4), 15 (2002) [Tech. Phys. **47**, 389 (2002)].
11. S. O. Shiryaeva, Zh. Tekh. Fiz. **73** (2), 19 (2003) [Tech. Phys. **48**, 152 (2003)].
12. A. R. Gaibov, S. O. Shiryaeva, A. I. Grigor'ev, and D. F. Belonozhko, Pis'ma Zh. Tekh. Fiz. **29** (4), 22 (2003) [Tech. Phys. Lett. **29**, 138 (2003)].
13. A.-H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973; Mir, Moscow, 1976).
14. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
15. I. P. Stakhanov, Zh. Tekh. Fiz. **44**, 1373 (1974) [Sov. Phys. Tech. Phys. **19**, 861 (1974)].
16. A. I. Grigor'ev, S. O. Shiryaeva, and V. A. Koromyslov, Zh. Tekh. Fiz. **68** (9), 1 (1998) [Tech. Phys. **43**, 1011 (1998)].
17. S. O. Shiryaeva, A. I. Grigor'ev, and D. F. Belonozhko, Zh. Tekh. Fiz. **69** (10), 34 (1999) [Tech. Phys. **44**, 1159 (1999)].

Translated by N. Mende

Evaluation of Plasma Parameters
from Measured Data
on the Passage of Shock Waves through a Plasma

V. V. Kuchinskiĭ

Research Institute of Hypersonic Systems, Leninets Holding Company, St. Petersburg, 196066 Russia
e-mail: ajax@comset.net

Abstract—Analytic formulas are derived that make it possible to estimate the plasma parameters from measured data on the characteristics of a shock wave that has passed through a plasma. Examples are given of how these formulas can be used to interpret the data from experimental studies of propagation of a shock wave through a gas-discharge plasma. The results obtained can serve as a starting point for a more detailed analysis. © 2003 MAIK "Nauka/Interperiodica".

INTRODUCTION

Among the first papers on the diagnostics of plasmas by shock or acoustic waves are [1, 2], in which an argon plasma at pressures from 2 to 30 torr was investigated by Langmuir probes. Sharp variations in the detected waveforms of the potential made it possible to estimate the wave velocity and to calculate the dependence of the discharge temperature on the experimental conditions.

Further development of experimental techniques increased the variety of tools for recording the velocity of a shock or an acoustic wave. The objective of the present paper is to elaborate simple methods for estimating the plasma parameters from the experimentally measured characteristics of a shock wave. Of course, more precise modern methods of calculation provide far more precise information about the plasma but, as a rule, they involve a considerable expenditure of computational resources. There are many cases in which simple estimates may be useful, because variations in the plasma parameters make it practically impossible to determine the temperature (or density) of the neutral plasma component with a high degree of accuracy. Section 1 presents the main formulas for calculating the plasma temperature from measured data on the passage of a shock wave through the spatial inhomogeneities that originate from the nonuniform temperature (and, accordingly, density) distribution at a constant pressure in an unperturbed region ahead of the shock front. Section 2 demonstrates how the theoretical results obtained can be used to interpret previous experimental data. This is most fully illustrated by processing the data of [3, 4]. Less accurate methods for processing the experimental information are described in the Appendix.

1. THEORETICAL BASIS
FOR THE METHOD

A detailed analytic treatment of the formation of a shock wave and its passage through a spatially inhomogeneous region was carried out by Sukhomlinov *et al.* [5] in the one-dimensional approximation. The results obtained in that paper yield the following set of equations:

$$\frac{\partial x(t)}{\partial t} = a(x(t)) + a_0 \frac{f(x(t))}{2},$$
$$\frac{\partial (tf(x(t)))}{\partial t} = \frac{a_0 f(x(t))}{2a(x(t))}, \tag{1}$$

where x is the spatial coordinate, t is the time, and the function $f(t)$ describes the initial shape of the shock wave perturbation.

When the spatial inhomogeneity is produced by a nonuniform temperature (or density) distribution at a constant pressure, the acoustic velocity at the point x is equal to $a(x) = a(T(x)) = \sqrt{\gamma R T(x)}$, where R is the universal gas constant, $a_0 = a(T_0) = \text{const}$, T_0 is the constant temperature outside the spatially inhomogeneous region, $T(x)$ is the shape of the temperature distribution inside the inhomogeneous region, and γ is the adiabatic index.

The solution to Eqs. (1) for a shock wave propagating through a homogeneous medium yields a number of simple formulas describing how the wave changes in time. The initial coordinate x_0 of the shock front is related to the initial time t_0 and the initial value of the Mach number M_0 by

$$x_0 = (2M_0 - 1)a_0 t_0. \tag{2}$$

The time dependence of the shock front coordinate,

$$x(t) = 2(M_0 - 1)a_0\sqrt{t_0 t} + a_0 t, \qquad (3)$$

gives the value of the Mach number at the time t:

$$M(t) = \frac{\partial x(t)}{\partial t}\frac{1}{a_0} = \frac{V_0(t)}{a_0}$$

$$= (M_0 - 1)\sqrt{\frac{t_0}{t}} + 1 = \frac{1}{2} + \frac{x(t)}{2a_0 t}, \qquad (4)$$

where V_0 is the shock front velocity and $M_0 = M(t_0)$.

From formula (3) we obtain the dependence of the time on the coordinate, $t^{(0)}(x)$:

$$t^{(0)}(x) = \frac{x_0}{a_0}\frac{(M_0 - 1)^2}{2M_0 - 1}\frac{1}{F(x, x_0, M_0)^2}, \qquad (5)$$

where the subscript 0 stands for the motion in a homogeneous medium and

$$F(x, x_0, M_0) = \frac{1}{\sqrt{1 + \dfrac{x}{x_0}\dfrac{(2M_0 - 1)}{(M_0 - 1)^2}} - 1}. \qquad (6)$$

Substituting expression (5) into formula (4), we arrive at the dependence of the velocity of a shock wave in a homogeneous medium on the coordinate x:

$$V^{(0)}(x, x_0, M_0, a_0) = a_0[1 + F(x, x_0, M_0)]. \qquad (7)$$

Of great practical interest is the case of a "rectangular" temperature profile such that $T(x) = T_0$ for $x_p < x \le x_p$ and $T(x) = T_0$ for $x \le x_p$ and $x > x_f$ (see Fig. 1). In this case, Eqs. (1) admit an exact solution, which yields the following expression for the shock wave velocity:

$$V(t, T) = \frac{dx(t, T)}{dt} = a_0$$

$$\times \begin{cases} 1 + (M_0 - 1)\sqrt{\dfrac{t_0}{t}} & \text{for } t \le t_p \\[2mm] \sqrt{\dfrac{T}{T_0}} + (M_0 - 1)\sqrt{\dfrac{t_0}{t}}\left(\dfrac{t}{t_p}\right)^{\Psi} & \text{for } t_p < t \le t_f(T) \\[2mm] 1 + (M_0 - 1)\sqrt{\dfrac{t_0}{t}}\left(\dfrac{t_f(T)}{t_p}\right)^{\Psi} & \text{for } t > t_f(T) \\[2mm] \Psi = \dfrac{1}{2}\sqrt{\dfrac{T_0}{T}} - \dfrac{1}{2}. \end{cases} \qquad (8)$$
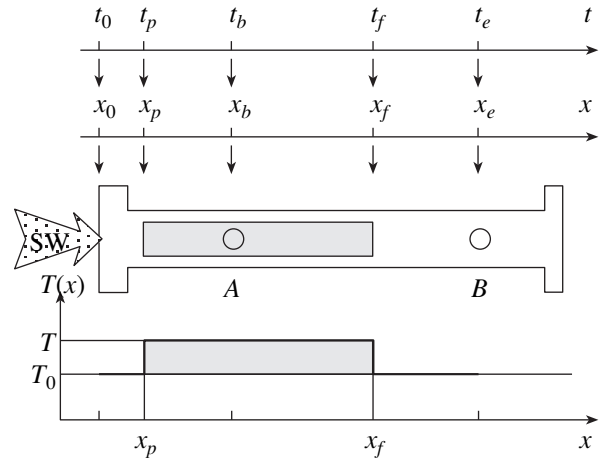
We integrate this expression to get the time depen-

Fig. 1. Choice of the coordinate system: SW is a shock wave propagating from left to right; x_0 is the initial coordinate of the shock front (the position of the membrane); x_p and x_f are the entrance and exit boundaries of the inhomogeneous region, respectively; and x_b and x_e are the coordinates of the points A and B at which a shock wave (SW) is recorded.

dence of the coordinate of the shock front:

$$x(t, T)$$

$$= \begin{cases} a_0[2(M_0 - 1)\sqrt{t_0}\sqrt{t} + t] & \text{for } t \le t_p \\[2mm] x_p + \sqrt{\gamma R T}\left[t - t_p + 2(M_0 - 1)\sqrt{t_0}\sqrt{t_p}\right. \\[2mm] \left. \times\left(\left(\dfrac{t}{t_p}\right)^{\Psi + \frac{1}{2}} - 1\right)\right] & \text{for } t_p < t \le t_f(T) \\[2mm] x_f + a_0\left[t - t_f(T) + 2(M_0 - 1)\sqrt{t_0}\right. \\[2mm] \left. \times\left(\dfrac{t_f(T)}{t_p}\right)^{\Psi}(\sqrt{t} - \sqrt{t_f(T)})\right] & \text{for } t > t_f(T). \end{cases} \qquad (9)$$

The inverse function $t(x, T)$ is found by solving Eq. (9) with respect to t. To within a relative error of less than 2×10^{-5}, the coordinate dependence of the time is given by the expression

$$t(x, T)$$

$$\approx \begin{cases} t^{(0)}(x) & \text{for } t \le t_p, \\[2mm] \dfrac{\beta(x, T)^{\Psi + 1/2}(\Psi - 1/2)\omega + 1 + \omega + (\Psi + 1/2)\eta(x)\omega}{1 + (\Psi + 1/2)\beta(x, T)^{\Psi - 1/2}\omega}t_p \\[2mm] \text{for } t_p < t \le t_f(T), \\[2mm] \left[\sqrt{\dfrac{x - x_f}{a_0} + (\sqrt{t_r(T)} + \sqrt{t_s(x_f, T)})^2} - \sqrt{t_r(T)}\right]^2 \\[2mm] \text{for } t > t_f(T), \end{cases} \qquad (10)$$

where

$$\omega = 2(M_0 - 1)\sqrt{\frac{t_0}{t_p}}, \quad t_r(T) = (M_0 - 1)^2 \left(\frac{t_f(T)}{t_p}\right)^{2\Psi} t_0,$$

$$\eta(x) = \frac{x - x_p}{a_0\sqrt{t_0 t_p}(M_0 - 1)},$$

$$\beta(x, T) = 1 + \frac{t_{be}^{(0)}(x_p, x, x_p, \sqrt{\gamma RT}, M_x)}{t_p}, \quad (11)$$

$$M_x = 1 + \sqrt{\frac{T_0}{T}} F(x_p, x_0, M_0),$$

$$t_{be}^{(0)} = (x_b, x_e, x_0, a_0, M)$$

$$= \frac{1}{a_0}\left[x_e - x_b - \frac{2x_0}{\Theta}\left(\sqrt{1 + \Theta\frac{x_e}{x_0}} - \sqrt{1 + \Theta\frac{x_b}{x_0}}\right)\right], \quad (12)$$

$$\Theta = \frac{2M - 1}{(M - 1)^2},$$

and $t_{be}^{(0)}$ is the time required for the wave with $M = M_0$ to pass from the point x_b to the point x_e in a homogeneous medium.

The time t_p at which a shock wave reaches the entrance boundary of a thermally inhomogeneous region is independent of temperature and is calculated from formula (5): $t_p = t^{(0)}(x_p)$. The time $t_f(T)$ at which the wave leaves the inhomogeneous region is obtained as a solution to Eq. (9), i.e., to the equation $x(t_f(T), T) = x_f$. With a high degree of accuracy, this time can be approximated by the expression $t_f(T) \approx t(x_f, T)$, which is obtained using the second of equalities (10). The time dependence of the shock wave velocity is given by formula (8), and the coordinate dependence of this velocity is calculated from formula (8) with the help of formula (10):

$$V(x, T) = V(t(x, T), T). \quad (13)$$

In the general case, the time required for the wave to pass the distance between the points at which it is recorded (i.e., the points x_b and x_e) is determined by the formula

$$t_{be}(T) = \int_{x_b}^{x_e} \frac{dx}{V(x, T)} = t(x_e, T) - t(x_b, T), \quad (14)$$

where, depending on circumstances, the velocity $V(x, T)$ is given by formulas (7), (8), and (13) or by approximate formulas (24) and (25) from the Appendix and the time $t(x, T)$ is given by formula (10). For a shock wave propagating in a homogeneous medium, the two groups of formulas for the velocity are equivalent, so that we have $t_{be}(T_0) = t_{be}^{(0)}$ (see formula (12)).

When the initial value of the Mach number is unknown, it can be determined by measuring the time t_{be} required for the wave to pass the distance between the points at which it is recorded in a homogeneous space (with a uniform temperature) and substituting the result obtained into formula (12):

$$M_0(t_{be})$$

$$= \left[1 - \frac{1}{\sqrt{1 + 4x_0\dfrac{x_b + x_e - \sqrt{(x_b + x_e - a_0 t_{be})^2 + 4x_b a_0 t_{be}}}{(x_e - x_b - a_0 t_{be})^2}}}\right]^{-1}. \quad (15)$$

Using formula (14), we can calculate the time as a function of a parameter set: $t_{be} = t_{be}(T, M_0, x_0, x_b, x_e)$. If the time of propagation of a shock wave, $(t_{be})_{\exp}$, is known from experiment, then, by solving the transcendental equation $t_{be}(T, M_0, x_0, x_b, x_e) = (t_{be})_{\exp}$ with respect to the temperature T, we determine the temperature (or the density $\rho/\rho_0 = T_0/T$) in a spatially inhomogeneous region.

When the wave characteristics are measured at points inside a thermally inhomogeneous region and the first measurement point x_b coincides with its entrance boundary (i.e., $x_b = x_p$, $t_b = t_p$, $x_e \le x_f$), the temperature can be determined with high accuracy:

$$T(t_{be}, M_p)$$

$$\approx T_0 \frac{\tau^4}{4\left[\tau - \eta_{be} + \sqrt{(\tau - \eta_{be})^2 - \dfrac{t_{be}}{\sqrt{t_x t_b}}\tau^2}\right]^2}, \quad (16)$$

where

$$\eta_{be} = \frac{x_e - x_p}{a_0\sqrt{t_x t_p}}, \quad \tau = \ln\left(1 + \frac{t_{be}}{t_p}\right),$$

$$t_x = \frac{x_p F(x_p, x_0, M_0)^2}{a_0[1 + 2F(x_p, x_0, M_0)]}.$$

The maximum temperature in a spatially inhomogeneous region can be estimated not only from the data on variations in the velocity of a shock wave or its propagation time but also from the change in the relative intensity of the measured signal (see Fig. 2 taken from [4] and analogous figures in [6]). A decrease in the intensity of the signal recorded by a photodetector is determined, on the one hand, by a decrease in the density in the region of elevated temperature (($\rho/\rho_0 = T_0/T$, $T > T_0$) and, on the other, by a reduction in the local Mach number. When the recorded parameter is the first

derivative of the density, the relative change in the signal intensity can be approximately described as

$$\frac{I_1(j)}{I_1(j_c)} \approx \frac{\partial \rho(j)/\partial x}{\partial \rho(j_c)/\partial x} \approx \frac{T_c}{T}\frac{J_1(M)}{J_1(M_c)},$$ (17)

$$J_1(M) \approx 1.24\{1 - \exp[-1.7(M-1)]\}(M-1)^{1.78}.$$

The expression for the dependence of the maximum absolute value of the first derivative of the density on the Mach number is an approximation of the general (but fairly involved) analytic formula presented in the Appendix.

When the recorded parameter is the second derivative of the density, the corresponding formulas have the form

$$\frac{I_2(j)}{I_2(j_c)} \approx \frac{\partial^2 \rho(j)/\partial x^2}{\partial^2 \rho(j_c)/\partial x^2} \approx \frac{T_c}{T}\frac{J_2(M)}{J_2(M_c)},$$ (18)

$$J_2(M) \approx 1.54\{1 - \exp[-1.7(M-1)]\}(M-1)^{3.07}.$$

In formulas (17) and (18), j_c is the current density at which the temperature T_c and Mach number M_c are known. For other values of the current density, the Mach number is a function of temperature, $M = M(T)$; at the measurement point x_e, it is calculated from formulas (7), (8), and (13) or from formulas (24) and (25): $M(T) = V(x_e, x_0, M_0, a_0, T)/a(T(x_e))$. In this case, the temperature in a thermally inhomogeneous region is estimated from the solution to the transcendental equation

$$\left[\frac{I_1(j)}{I_1(j_c)}\right]_{exp} = \frac{I_1(j)}{I_1(j_c)} \quad \text{or} \quad \left[\frac{I_2(j)}{I_2(j_c)}\right]_{exp} = \frac{I_2(j)}{I_2(j_c)},$$ (19)

where $[I_1(j)/I_1(j_c)]_{exp}$ or $[I_2(j)/I_2(j_c)]_{exp}$ are the experimentally measured ratio of the signal amplitudes (see, e.g., Fig. 2) and the ratios $I_1(j)/I_1(j_c)$ or $I_2(j)/I_2(j_c)$ are calculated from formulas (17) and (18) (depending on the method by which the signal is detected) and from the formula $M(T) = V(x_e, x_0, M_0, a_0, T)/a(T(x_e))$.

In many cases, it is expedient to interpret experimental data using the following formula, which was proposed by Sukhomlinov et al. [7] and describes the dependence of the temperature T of a gas discharge on the current density, pressure, and the ratio E/p (characterizing the sort of gas):

$$T = cT_0\varphi, \quad \varphi = \frac{1}{\sqrt{3}}\left[1 + \left(\frac{1-q}{1+q}\right)^{0.4}\right]\left(\frac{1+q}{1-q}\right)^{0.2},$$

$$q = \sqrt{1 - \left(\frac{j_0}{j}\right)^2}, \quad j_0 = \frac{21.383}{\left(\frac{E}{p}\right)p}.$$ (20)

Here, cT_0 is the temperature of the wall surrounding the discharge, the electric field strength E is in V/cm, the current density j is in mA/cm^2, the pressure p is in torr,
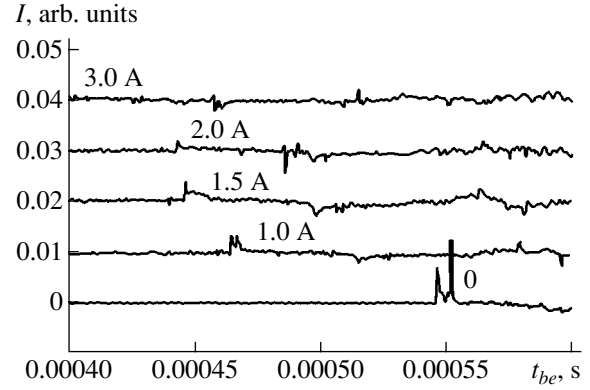

Fig. 2. Measurement of the effect of the plasma on the propagation of a shock wave at a pressure of 1.6 torr (the time of propagation between the points at which the wave is recorded, according to the data of [3, 4]).

and the numerical constant depends on the choice of the units of measure.

The constant $c \geq 1$ can be called the coefficient of heating of the wall with respect to the surrounding medium, or simply the "heating coefficient." Formula (20) is valid only for $j \geq j_0$. The following generalized temperature profile can be proposed that is valid over the entire range $j \geq 0$:

$$T = \begin{cases} T_0\left[\left(1 - \frac{42c}{25\sqrt{3}}\right)\left(\frac{j}{j_0}\right)^2 - 2\left(1 - \frac{46c}{25\sqrt{3}}\right)\left(\frac{j}{j_0}\right) + 1\right] \\ \text{for } j \leq j_0, \\ cT_0\varphi \quad \text{for } j > j_0. \end{cases}$$ (21)

If the wall temperature is determined by a mean of the temperature T_0 and the plasma temperature, i.e., by $T = (T + T_0)c_b\varphi$ (where the coefficient c_b is about 1/2), then, instead of formula (21), we obtain

$$T$$

$$= \begin{cases} T_0\left[\left(\dfrac{3 - (142/25)c_b\sqrt{3} + 8c_b^2}{(2c_b - \sqrt{3})^2}\right)\left(\dfrac{j}{j_0}\right)^2 \\ \quad - 2\left(\dfrac{3 - (146/25)c_b\sqrt{3} + 8c_b^2}{(2c_b - \sqrt{3})^2}\right)\left(\dfrac{j}{j_0}\right) + 1\right] \text{for } j \leq j_0, \\ c_bT_0\varphi \quad \text{for } j > j_0. \end{cases}$$ (22)

These formulas allow one to estimate the plasma temperature and also, in the case under consideration, can provide additional tools for checking the temperature values obtained from the experimental data.

Below, examples will be presented of how the above methods for determining the temperature can be applied in practice (Figs. 3–6). When the pressures

inside and outside the thermally inhomogeneous region are the same, these methods make it possible to completely determine the main parameters inside this region, provided that the pressure P_1 and the temperature T_0 outside it are known.

2. DETERMINATION OF THE PLASMA TEMPERATURE FROM THE MEASURED SHOCK WAVE CHARACTERISTICS

Papers [3, 4] contain fairly complete information about the experimental conditions and the results obtained. In those experiments, the propagation time of a shock wave between two measurement points was recorded by two lasers. The experiments were carried out with air at pressures of 1.6 and 10 torr. A shock wave passed through a gas-discharge plasma in a direction transverse to the applied electric field. The discharges could occur in different regions of the working chamber. The notation adopted for describing the spatial temperature distribution over the thermally inhomogeneous discharge region is given in Fig. 1.

The characteristic waveforms [4] from which the propagation time of a shock wave between the measurement points can be determined are presented in Fig. 2. Substituting the propagation time $t = 0.00055$ s in a homogeneous (nonionized) medium into formula (15), we determine the Mach number just at the membrane, $M_0 = 2.88$. This value coincides with the corresponding data of [3, 4]. Then, for the purpose of checking the results, we can calculate the propagation time of a shock wave at nonzero currents from formulas (10) and (14), in which the temperature is determined from formulas (21) and (22). Since the temperature of the discharge chamber wall was not measured in [3, 4], the coefficient c in formula (21) can be adjusted by estimating the power fed into the discharge. In Fig. 3b, the solid curve shows the temperature profile calculated from formula (21) with $c = 1.7$, and the dotted curves are obtained from the same formula but with $c = 1.6$ (lower curve) and $c = 1.8$ (upper curve). The dashed curve is calculated from formula (22) with $c_b = 1/2$. The squares show the results obtained from the experimentally measured relative signal amplitudes (this method for estimating T was described in Section 1). It may be said that formula (21) describes the experimental situation in question more adequately. The dependence of the temperature on the current density is calculated by solving the equation $t_{be}(T, M_0, x_0, x_b, x_e) = (t_{be})_{\exp}$. The estimates of the temperature that are obtained from the relative signal amplitudes are apparently less accurate
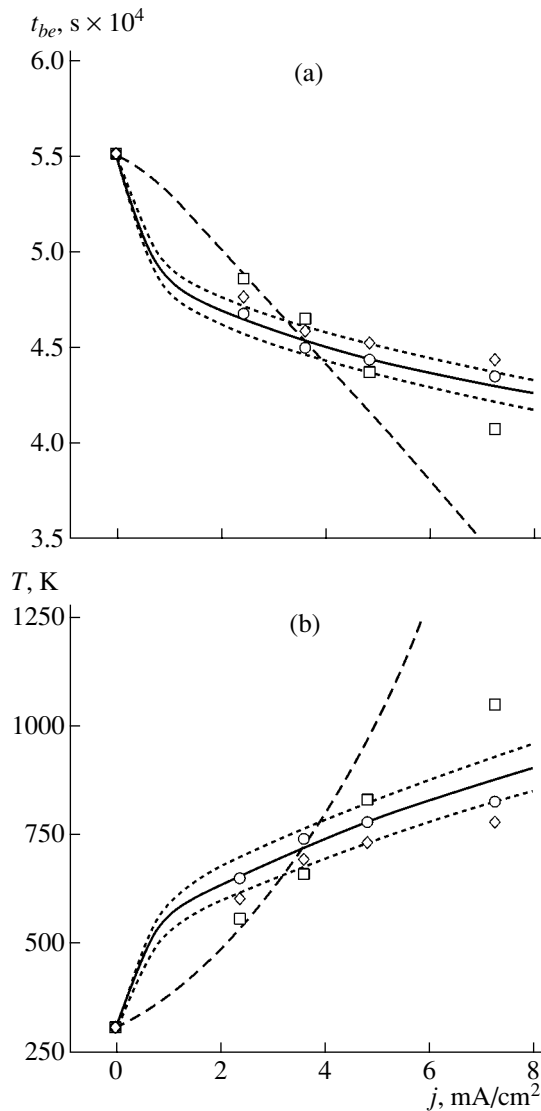

Fig. 3. (a) The time t required for a shock wave with the initial Mach number $M_0 = M(x_0) = 2.88$ to pass the distance between the points at which it is recorded and (b) the temperature T in the case of a rectangular temperature profile as functions of the current density j. The notation in plot (a) is as follows: The circles show the experimental data of [4] (see Fig. 2). All the curves were computed from formula (14). The solid curve was obtained by using the temperature calculated from formula (21) with $c = 1.7$, and the upper and lower dotted curves were obtained by using the temperatures calculated from formula (21) with $c = 1.6$ and $c = 1.8$, respectively. The dashed curve was computed from formula (14) in which the temperature was calculated from formula (22). The squares show the results calculated from formula (14) in which the temperature was determined from the relative signal amplitude by solving Eq. (19). The diamonds show the results calculated from formula (14) in which the temperature was estimated from formula (26) with the use of the experimental data on the propagation time of a shock wave [4]. The notation in plot (b) is as follows: The solid curve is the temperature profile calculated from formula (21) with $c = 1.7$, and the upper and lower dotted curves were calculated from the same formula with $c = 1.8$ and $c = 1.6$, respectively. The dashed curve was calculated from formula (22). The circles show the temperature calculated from the experimental data of [4] by solving the equation $t_{be}(T, M_0, x_0, x_b, x_e) = (t_{be})_{\exp}$. The squares show the temperature determined from the relative signal amplitude by solving Eq. (19). The diamonds show the temperature calculated from formula (26) with the use of the experimental data on the propagation time of a shock wave [4].
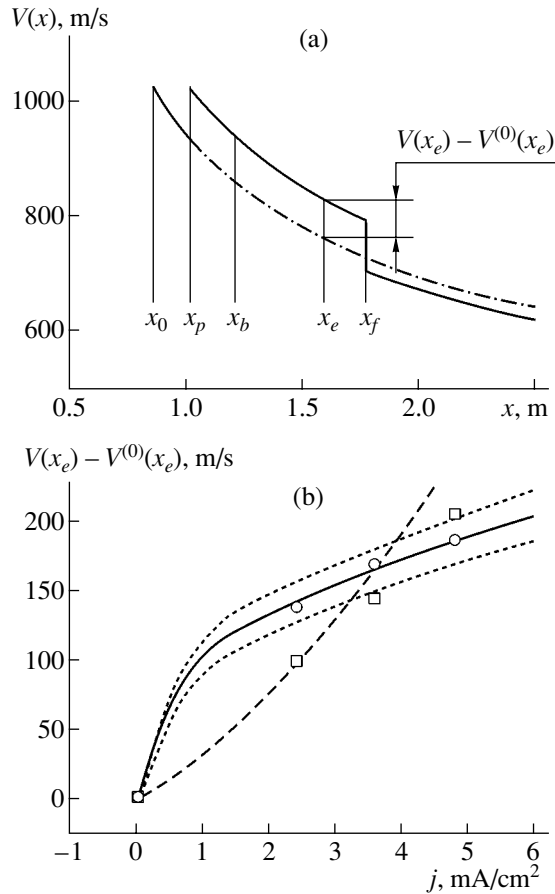
Fig. 4. (a) Profile of the shock wave velocity along the spatial coordinate and (b) increment in the velocity caused by the spatial inhomogeneity. The notation in plot (a) is as follows: The solid curve is the velocity $V(x)$ of a shock wave with the initial Mach number $M_0 = 2.88$. The dashed-and-dotted curve is the velocity $V^{(0)}(x)$ calculated from formula (7) for a shock wave that has the same initial Mach number $M_0 = 2.88$ but moves in a homogeneous medium. Plot (b) displays the difference (at the observation point x_e) between the velocities shown in plot (a). The calculations were carried out on the basis of formula (14) in which the temperature (Fig. 3b) was determined from formula (21) with $c = 1.7$ (solid curve), from the same formula but with $c = 1.8$ (upper dotted curve) and $c = 1.6$ (lower dotted curve), from formula (22) (dashed curve), from the experimental data obtained in [4] on the propagation time at the pressure $p = 1.6$ torr (circles), and from the measured signal amplitude (squares).
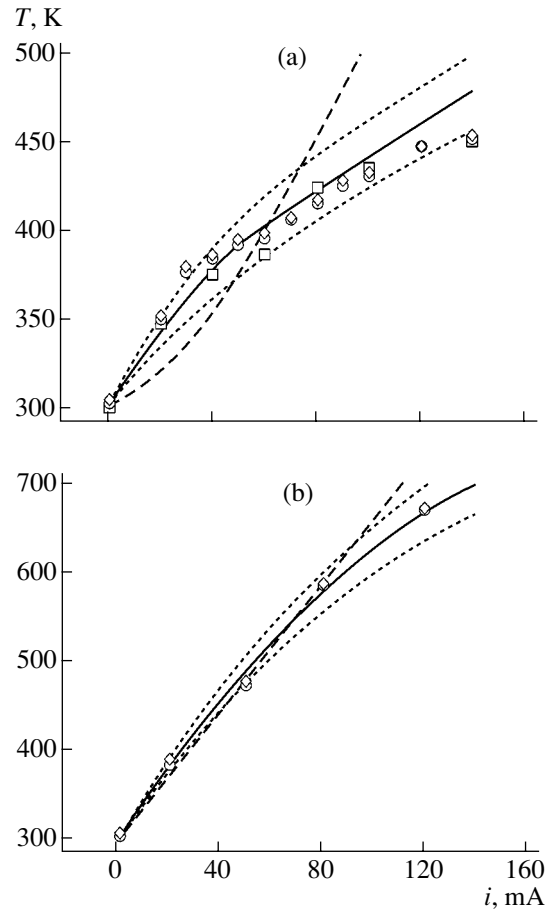
Fig. 5. Dependence of the temperature on the current for a shock wave propagating through a discharge in (a) argon at the pressure $p = 30$ torr and (b) nitrogen at the pressure $p = 10$ torr. The circles show the temperature values determined by solving the equation $t_p(T, M_0, x_0, x_b, x_e) = (t_p)_{exp}$ or calculated from formula (26) with the use of the experimental data obtained in [6] on the propagation time of a shock wave. The squares show the temperature values determined from the relative signal amplitude [6] by solving Eq. (19). In plot (a), the solid curve shows the temperature calculated from formula (21) with $c = 1.15$, and the dotted curves were obtained from the same formula but with $c = 1.2$ (upper curve) and $c = 1.1$ (lower curve). The dashed curve was calculated from formula (22). In plot (b), the solid curve shows the temperature calculated from formula (21) with $c = 2.1$, and the dotted curves were obtained from the same formula but with $c = 2.2$ (upper curve) and $c = 2.0$ (lower curve). The dashed curve was calculated from formula (22).

than those obtained from the measured propagation times.

The solid curve in Fig. 4a illustrates the results of calculations of the shock wave propagation velocity as a function of the spatial coordinate from formula (13) at $T = 473$ K. The dashed-and-dotted curve shows the propagation velocity calculated from the same formula but at $T = T_0 = 300$ K, i.e., in the absence of thermal inhomogeneity. The difference between these curves is the increment in the velocity at the observation point x_e

caused by the effect of the thermal inhomogeneity (Fig. 4b).

The data obtained by Ganguly and Bletzinger [6] on the propagation of a shock wave in an argon and a nitrogen plasma are illustrated in Figs. 5a and 5b, respectively. In that paper, the propagation time was measured from the waveforms that are analogous to those given in Fig. 2. It is obvious that, in this case, the estimates of the temperature from the propagation time of a shock wave have almost the same accuracy as those obtained from the signal amplitude.
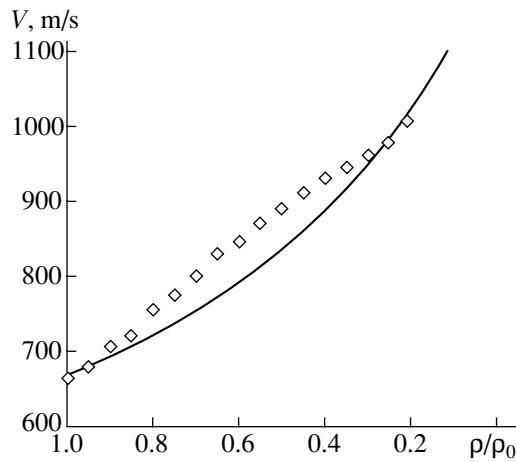
Fig. 6. Shock wave velocity as a function of the relative density. The diamonds show the experimental data of [8]. The solid curve was calculated from formula (24) in which the initial Mach number was determined from formula (33), the pressure being $p = 60$ torr. The density increases linearly according to the law $\rho_i(x)/\rho_{10} \approx 0.2 + 0.8(x - x_p)/(x_f - x_p)$ [8].

Piskareva and Shugaev [8] carried out experiments on the propagation of shock waves with different initial velocities through an inhomogeneous region in which the density increases according to a linear law, $\rho_1(x)/\rho_{10} \approx 0.2 + 0.8(x - x_p)/(x_f - x_p)$ (in [8], this law was determined experimentally). They measured the distributions of the shock wave velocity along the spatial coordinate. The maximum temperature estimated by the method of two velocities described in the Appendix turns out to be $T = 1404$ K, which agrees well with the estimate $T \approx 1370$ K obtained in [8]. Figure 6 compares the calculated shock wave velocity as a function of the distance from the heat source with the experimental results of [8]. Good agreement between the calculated and experimental results raises the hope that this method will also yield fairly reliable estimates of the temperature.

CONCLUSIONS

In this paper, formulas have been derived for estimating plasma parameters by several independent methods from experimental data on the characteristics of shock waves passing through a thermally inhomogeneous region. These formulas allow one to calculate the shock wave velocity and the time required for the wave to pass the distance between the points at which it is recorded. A comparison of the results obtained by the methods proposed here with the experimental data gives reason to hope that the methods can be successfully used to estimate the plasma parameters in different experimental situations.

For a homogeneous medium, the function $f(t)$ has the form

$$f(t) = \frac{2(M_0 - 1)\sqrt{t_0}}{\sqrt{t}}. \tag{23}$$

As a first approximation, we can substitute this expression for $f(t)$ into the first of Eqs. (1) and switch from the dependence on the time t to the dependence on the coordinate x according to formula (5). Then, we can derive the following simple approximate formula for the velocity of a shock wave propagating in a medium with a spatial inhomogeneity:

$$V(x, x_0, M_0, a_0, T) \approx a(T(x)) + a_0 F(x, x_0, M_0), \tag{24}$$

where T is the maximum temperature in a thermally inhomogeneous region and $M_0 = M(x_0)$.

In the next approximation, the shock wave propagation velocity is described by a more involved formula:

$$V(x, x_0, M_0, a_0, T) \approx a(T(x)) + a_0 F(x)$$

$$\times \begin{cases} 1 & \text{for} \quad x \le x_p, \\ \sqrt{\dfrac{1 + F(x_p)\left(\dfrac{T_0}{T}\right)^{1/3} + \left[1 - \left(\dfrac{T_0}{T}\right)^{1/3}\right]\dfrac{F(x)^2}{F(x_p)}}{1 + F(x_p)}} & \text{for } x > x_p, \end{cases} \tag{25}$$

$$F(x) = F(x, x_0, M_0), \quad F(x_p) = F(x_p, x_0, M_0) = \text{const}.$$

A comparison with the numerical solution to the exact equations shows that formula (25) is valid for different temperature profiles.

For a rectangular temperature profile (Fig. 1), the simplest way to estimate the temperature in a thermally inhomogeneous region is to replace the dependence of the shock wave velocity on the coordinate x in the integrand in formula (14) with the value of the velocity in the middle of the interval between the measurement points x_b and x_e (such that $x_b \ge x_p$, $x_e \le x_f$):

$$T \approx \frac{1}{\gamma R}\left[\frac{x_e - x_b}{(t_{be})_{\exp}} - a_0 F\left(\frac{x_e + x_b}{2}, x_0, M_0\right)\right]^2. \tag{26}$$

If the velocity of a shock wave is maximum just after the wave enters the thermally inhomogeneous region, then the temperature in this region can be estimated from the measured spatial distributions of the velocities of two shock waves with different initial Mach numbers M_{01} and M_{02}. In this case, the maximum velocity is given by the formula

$$V_{\max} = a_0\left(M_p - 1 + \sqrt{\frac{T}{T_0}}\right), \tag{27}$$

where $a_0 M_p$ is the shock wave velocity just before the wave enters the thermally inhomogeneous region, M_p is the Mach number corresponding to the maximum

velocity, and T is the maximum temperature in the inhomogeneous region.

Using the measured distributions of the velocities of the two shock waves, we can construct the following set of four equations:

$$V_{1\max} = \left(M_{1p} - 1 + \sqrt{\frac{T}{T_0}}\right)a_0,$$

$$V_{2\max} = \left(M_{2p} - 1 + \sqrt{\frac{T}{T_0}}\right)a_0,$$

$$V_{10} = V^{(0)}(x_m, x_0, M_{1p}, a_0),$$

$$V_{20} = V^{(0)}(x_m, x_0, M_{2p}, a_0),$$

(28)

the left-hand sides of which are the experimentally measured maximum velocities $V_{1\max}$ and $V_{2\max}$ and the velocities V_{10} and V_{20} measured at the point x_m that lies behind the spatially inhomogeneous region.

The quantity x_0, which determines the choice of the coordinate system, is an unknown parameter. The left-hand sides of the last two equations are calculated from formula (7). The set of equations contains four unknown quantities: M_{1p}, M_{2p}, T, and x_m/x_0. In solving the set of equations, we arrive at the following cubic equation for the quantity M_{1p}:

$$A_3 M_{1p}^3 + A_2 M_{1p}^2 + A_1 M_{1p} + A_0 = 0,$$

$$A_3 = 2g - 2, \quad A_2 = 2gq - 5g - 4q + 5,$$

$$A_2 = -4gq + 4g - 2q^2 + 6q - 4,$$

(29)

$$A_0 = 2gq + q^2 - 2q - g + 1,$$

$$g = \frac{(2V_{10} - a_0)(V_{20} - a_0)^2}{(V_{10} - a_0)^2(2V_{20} - a_0)}, \quad q = \frac{V_{2\max} - V_{1\max}}{a_0}.$$

The remaining unknown quantities are expressed in terms of the largest root M_{1p} of this equation:

$$T = \frac{[V_{1\max} - a_0(M_{1p} - 1)]^2}{\gamma R}, \quad M_{2p} = q + M_{1p},$$

$$x_0 = \frac{(x_m - x_0)_{\exp}(V_{10} - a_0)^2(2M_{1p} - 1)}{(a_0 M_{1p} - V_{10})(2M_{1p}V_{10} - V_{10} - a_0 M_{1p})},$$

(30)

where $(x_m - x_0)_{\exp}$ are the experimentally measured distance from the source of the shock wave to the point x_m, at which the velocities V_{10} and V_{20} were measured.

The calculated value of x_0 (Fig. 1) and the known value of $(x_m - x_0)_{\exp}$ give all the coordinates required for calculations, thereby making it possible to calculate the velocity distribution from formulas (8), (13), and (24) or (25) for the purpose of checking the results obtained.

If it is technically impossible to measure the pressure, then, by recording the time t_{be} required for a shock

wave to pass the distance between two points that lie outside of the spatially inhomogeneous region, one can use formula (15) to determine the Mach number at the exit from the shock tube, $M_0 = M_0(t_{be})$, and exploit the familiar formulas to calculate the pressure P_1 in the working volume from the pressure P_4 and temperature T_4 within the high-pressure chamber in the shock tube:

$$P_1 = \frac{P_4}{z}\left\{1 - \frac{(\gamma_4 - 1)(z - 1)\sqrt{\frac{\gamma T_0}{\gamma_4 T_4}}}{\sqrt{2\gamma[2\gamma + (\gamma + 1)(z - 1)]}}\right\}^{\frac{2\gamma_4}{\gamma_4 - 1}},$$

$$z = \frac{P_2}{P_1} = \frac{(2M_0 - 1)\gamma + 1}{\gamma + 1},$$

(31)

where γ_4 is the adiabatic index within the high-pressure chamber.

The maximum value of the first derivative of the density can be estimated by using the following formula, which was obtained from the experimental data of [9]:

$$\left.\frac{\partial\rho(x, M)}{\partial x}\right|_{x = x_{\max}}$$

$$\approx \frac{(1 - y)^2[(1 - 2y)(\sqrt{y^2 - y + 1} - 1) - 2y^2]}{(2y - 1 + \sqrt{y^2 - y + 1})^3}M,$$

(32)

$$y = \frac{(\gamma - 1)M^2 + 2}{(\gamma + 1)M^2}.$$

This formula is fairly well approximated by formula (17).

The initial Mach number of a shock wave propagating in a homogeneous medium can be estimated from the known velocity V_m (measured at the point x_m) by solving Eq. (7) with respect to M_0:

$$M_0 = B_m + \sqrt{(B_m - 1)B_m}$$

$$\left(B_m = \frac{x_m}{x_0}\frac{(V_m - a_0)^2}{(2V_m - a_0)a_0} + 1\right).$$

(33)

This formula can also be used to estimate the initial Mach number when the point x_m lies outside of the spatially inhomogeneous region, because the shock wave velocity in this region is close to that in the absence of inhomogeneity (see the region $x > x_f$ in Fig. 4a).

ACKNOWLEDGMENTS

REFERENCES

1. K. W. Gentle and U. Ingard, Appl. Phys. Lett. **5** (5), 105 (1964).
2. K. W. Gentle, U. Ingard, and G. Bekefi, Nature **23**, 1369 (1964).
3. D. M. Van Wie, A. L. Wesner, and L. R. Gauthier, in *Proceedings of the 3rd Workshop on Weakly Ionized Gases, Norfolk, 1999*, Paper AIAA-99-4824.
4. D. M. Van Wie and L. R. Gauthier, in *Proceedings of the 2nd Symposium on Thermochemical Processes in Plasma Aerodynamics, St. Petersburg, 2001*.
5. V. S. Sukhomlinov, V. Y. Kolosov, V. A. Sheverev, *et al.*, in *Proceedings of the 9th International Conference on Space Plane and Hypersonic Systems and Technologies, Norfolk, 1999*, Paper AIAA-99-4943.
6. B. N. Ganguly and P. Bletzinger, in *Proceedings of the Workshop on Weakly Ionized Gases, Colorado, 1997*, Vol. 2, p. HH1.
7. V. S. Sukhomlinov, V. Y. Kolosov, V. A. Sheverev, *et al.*, J. Fluid. Mech. **473**, 245 (2000).
8. M. V. Piskareva and F. V. Shugaev, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **19** (3), 11 (1978).
9. V. Y. Golyatin, A. L. Kuranov, V. V. Kuchinsky, *et al.*, in *Proceedings of the 32nd AIAA Conference on Plasmadynamics and Lasers Conference and 4th Weakly Ionized Gases Workshop, Anaheim, 2001*, Paper AIAA-2001-3055.

Translated by O. Khadin

GASES
AND LIQUIDS

On the Calculation of the Translational Mode Amplitude for a Drop Nonlinearly Vibrating in an Environment

S. O. Shiryaeva, A. I. Grigor'ev, V. A. Koromyslov, and A. N. Zharov

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

e-mail: shir@uniyar.ac.ru

Received February 20, 2003

Abstract—The second-order amplitudes of the capillary vibration modes of a drop of an ideal incompressible liquid placed in an incompressible ideal medium are calculated. The approximation is quadratic in initial multimode deformation of the equilibrium spherical shape caused by nonlinear interaction. The mathematical statement of the problem is such that the immobility condition for the center-of-mass of the drop is met automatically. When the translational mode amplitude is calculated, a set of hydrodynamic boundary conditions at the interface, rather than the condition of center-of-mass immobility (which is usually applied for simplicity in the problems of drops vibration in a vacuum), should be used. © 2003 MAIK "Nauka/Interperiodica".

(1) The vibration and stability of charged drops and bubbles in a liquid is of interest in various areas of science and technology (see, e.g., [1–3] and Refs. cited therein). This problem has been studied in many experimental and theoretical works using both the linear [1–3] and nonlinear [4–10] approximations in vibration amplitude. The study of nonlinearly vibrating drops and bubbles has only recently begun, and the techniques of solving such problems have not yet become commonly accepted. Many particular cases have been covered superficially, which sometimes causes errors. For example, this applies to the excitation of the translational mode in a nonlinearly vibrating drop, which appears in second- and third-order calculations [4, 8, 9, 11, 12]. The very fact of translational mode excitation in a drop of an incompressible liquid nonlinearly vibrating in a vacuum follows from the requirement that the center-of-mass of the drop be stationary. When the spectrum of modes specifying the drop initial deformation contains two or more modes with sequentially increasing numbers, the center-of-mass immobility leads to the appearance of the translational mode among the modes excited by nonlinear interaction [9]. In other words, the excitation of the translational mode compensates for the displacement of the center-of-mass, which results from a mass distribution that is asymmetric about the center of an equilibrium spherical drop when the spectrum of the initial deformation contains modes with sequential numbers. The time dependence of the translational mode amplitude is periodic, causing the drop vibrating in a gaseous atmosphere to emit dipole acoustic radiation [9, 13]. If the drop is charged, it becomes the source of dipole electromagnetic radiation [9, 14].

A second-order analytical expression for the translational mode amplitude in a drop nonlinearly vibrating in a vacuum can be derived from both the condition of center-of-mass immobility and a set of hydrodynamic boundary conditions on the free surface of the drop. In both cases, its form is the same [9]. The situation changes if the drop vibrates in an environment (or a bubble vibrates in a liquid): in this case, analytical expressions for the translational mode amplitude that are obtained from the condition of center-of-mass immobility and a set of boundary conditions at the interface are different at first glance [11]. It should be noted, however, that Benjamin and Ellis [11] applied the condition of center-of-mass immobility incorrectly. Nevertheless, they drew the general conclusion that the drop (bubble) moves translationally with a fixed velocity as a result of surface vibrations that are excited because of energy transfer from surface modes to the translational mode, whose amplitude contains a time-independent term. In combination with the misinterpretation of the experimental data for cavitation [15], this conclusion provoked another theoretical work by Feng [12], who, using an incorrect transition to the noninertial coordinate system, obtained an expression for the velocity of the translational motion of a bubble in a liquid in the absence of external forces (i.e., when the motion is caused only by surface vibrations). Therefore, the problem of applying the condition of center-of-mass immobility to calculating the nonlinear vibrations of incompressible liquid drops in an incompressible ideal environment in the case of a multimode initial deformation seems to be topical.

Kornfeld and Suvorov [15] observed the formation, motion, and cavitational disappearance of microbubbles in a liquid and near a metallic object vibrating at 7.5 kHz. The number of bubbles was rather large: they even clouded the object. Most of the bubbles formed and collapsed in the neighborhood of the object, thus causing its cavitational erosion. However, some of them suddenly started to move rapidly and randomly. Based

on this observation, the authors of [11, 12] put forward the idea that nonlinear vibrations cause the drift of the bubbles. We believe that the interpretation [11, 12] of the observations [15] is far from correct. It is obvious that any complex random motion of individual bubbles may be provided if the flow field of a liquid in the vicinity of a vibrating metallic object is superimposed on the velocity fields of randomly oriented intense hydrodynamic flows near cavitation bubbles [16] in the presence of gravitational and buoyancy force fields. As far as we know, no directed motion of bubbles in a quiescent liquid the only reason for which is surface vibrations in the absence of directed applied forces has been detected in experiments.

(2) Let a charged spherical drop of an ideal incompressible conducting liquid with a mass density ρ_1 be placed in an ideal incompressible dielectric liquid environment with a permittivity ε_* and mass density ρ_2. The surface tension coefficient at the interface is σ, and the total charge and radius of the drop are Q and R, respectively.

Consider the capillary vibrations of the interface that are induced by a small initial perturbation of its equilibrium spherical shape. We consider only axisymmetric distortions of the interface and write a related equation in the spherical coordinate system with the origin at the center-of-mass of the drop:

$$F(r, \Theta, t) \equiv r - r(\Theta, t) \equiv r - R[1 + \xi(\Theta, t)],$$
$$(|\xi|/R) \ll 1, \tag{1}$$

where $\xi(\Theta, t)$ is a dimensionless function describing the spherical surface deformation that is related to surface vibrations.

Since the drop vibration amplitudes are small, we may assume that the motion of both media is potential. The potentials of the liquid velocity fields are $\psi_1(\mathbf{r}, t)$ and $\psi_2(\mathbf{r}, t)$ inside and outside the drop, respectively.

The conductivity of the drop is taken to be sufficiently high in order that the characteristic time of charge redistribution over its surface be much shorter than the characteristic hydrodynamic time scales of the problem. Then, an electric field near the drop may be considered to be electrostatic with a potential Φ at any time instant.

The equations describing liquid motions in the system are

$$\Delta\psi_1 = 0 \quad (0 \leq r < r(\Theta, t));$$
$$\Delta\psi_2 = 0; \quad \Delta\Phi = 0 \quad (r > r(\Theta, t)) \tag{2}$$

with the conditions at the interface (see (1))

$$\frac{\partial\psi_1}{\partial n} = \frac{\partial\psi_2}{\partial n};$$

$$\frac{\partial F}{\partial t} + \boldsymbol{\nabla}\psi_1 \cdot \boldsymbol{\nabla}F = 0;$$

$$P_1 - P_2 - \rho_1\frac{\partial\psi_1}{\partial t} - \frac{\rho_1}{2}(\boldsymbol{\nabla}\psi_1)^2 + \rho_2\frac{\partial\psi_2}{\partial t}$$
$$+ \frac{\rho_2}{2}(\boldsymbol{\nabla}\psi_2)^2 + \frac{\varepsilon_*}{8\pi}(\boldsymbol{\nabla}\Phi)^2 = \sigma\mathrm{div}\,\mathbf{n}; \tag{3}$$

$$\Phi(\Theta, t) = \Phi_S(t);$$

$$-\frac{\varepsilon_*}{4\pi}\oint_S(\mathbf{n}\cdot\boldsymbol{\nabla}\Phi)dS = Q, \quad S = \begin{cases} r = R[1 + \xi(\Theta, t)] \\ 0 \leq \Theta \leq \pi \\ 0 \leq \phi \leq 2\pi. \end{cases}$$

Here, P_1 and P_2 are the pressures inside and outside the drop in the equilibrium state, $\mathbf{n}$ is the unit vector of the outer normal (directed toward the environment) to interface (1), $\Phi_S(t)$ is the constant value of the electrostatic potential $\Phi(\mathbf{r}, t)$ over the interface, and ∇ is the Laplacian.

The initial conditions are an expression for the initial diffusion of the interface and the zero initial velocity of the diffusion:

$$\xi(\Theta, t = 0) = \varepsilon\sum_{i \in \Xi} h_i P_i(\cos\Theta) + \xi_0 P_0(\cos\Theta) + \xi_1 P_1(\cos\Theta);$$
$$\frac{\partial\xi(\Theta, t = 0)}{\partial t} = 0. \tag{4}$$

Here, ε is the perturbation amplitude, which is a small parameter of the problem; $P_i(\cos\Theta)$ is the ith-order Legendre polynomial; h_i is the partial contribution of an ith vibration mode to the initial perturbation:

$$\sum_{i \in \Xi} h_i = 1,$$

ξ_0 and ξ_1 are constants defined by the constancy condition for the volume of the drop (and environment) with the vibrating interface,

$$\int_\Omega\int_0^{r(\Theta, t)} r^2 dr d\Omega = \frac{4}{3}\pi R^3; \quad \begin{aligned} &d\Omega \equiv \sin\Theta d\Theta d\varphi; \\ &\Omega = \{0 \leq \Theta \leq \pi; 0 \leq \varphi \geq 2\pi\} \end{aligned} \tag{5}$$

and the immobility condition for the center-of-mass of the entire system,

$$\frac{\rho_1\int_\Omega\int_0^{r(\Theta, t)} \mathbf{r}\cdot r^2 dr d\Omega + \rho_2\int_\Omega\int_{r(\Theta, t)}^{L} \mathbf{r}\cdot r^2 dr d\Omega}{\rho_1\int_\Omega\int_0^{r(\Theta, t)} r^2 dr d\Omega + \rho_2\int_\Omega\int_{r(\Theta, t)}^{L} r^2 dr d\Omega} = 0. \tag{6}$$

Conditions (5) and (6) must be met at any time instant, including at zero time. In (6), L is the characteristic linear size of the environmental space, $L \gg R$ (the environment occupies a relatively large volume, which

is infinitely large in the physical, rather than mathematical, sense).

(3) In the nonlinear statement, problem (2)–(4) may be solved by the method of many scales, as for a drop in a vacuum [4, 7–10]. Such an approach makes it possible to determine the function $\xi(\Theta, t)$, which is represented as the expansion in Legendre polynomials and describes the time evolution of the interface:

$$\xi(\Theta, t) = \sum_{n=0}^{\infty} [\varepsilon M_n^{(1)}(t) + \varepsilon^2 M_n^{(2)}(t) + O(\xi^3)] P_n(\cos\Theta). \quad (7)$$

In the problem of drop surface vibration in a vacuum ($\rho_2 = 0$), conditions (5) and (6) impose additional restrictions on the amplitudes of the zero (volume) and first (translational) modes in expansion (7), respectively, and these restrictions accord with the set of equations (2)–(4). For example, the expression for the translational mode amplitude that was derived from the condition of center-of-mass immobility coincided with that obtained from the set of boundary conditions. For the case of the drop placed in the environment, the role of condition (5) remains the same (since both media are assumed to be incompressible), whereas the application of condition (6) needs more detailed analysis.

First, we note that, projecting the integral of the vector function $\iint \mathbf{r} \cdot r^2 dr d\Omega$ onto the unit vectors of the Cartesian coordinate system, we obtain the equivalent system of three scalar integrals

$$\iint r^3 \sin\Theta\cos\varphi\, dr d\Omega; \quad \iint r^3 \sin\Theta\sin\varphi\, dr d\Omega;$$
$$\iint r^3 \cos\Theta\, dr d\Omega.$$

Combining these integrals, we can easily recast this system in compact form:

$$\iint r^3 Y_1^m(\Theta, \varphi)\, dr d\Omega \quad (m = -1; 0; 1),$$

where $Y_1^{\pm1}(\Theta, \varphi) \sim \sin\Theta \exp(\pm i\varphi)$ and $Y_1^0(\Theta, \varphi) \sim \cos\Theta$ are spherical functions.

In view of the above, we write condition (6) of center-of-mass immobility for the drop in the environment as

$$\frac{\int_\Omega \left[\rho_1 \int_0^{r(\Theta, t)} r^3 dr + \rho_2 \int_{r(\Theta, t)}^L r^3 dr\right] Y_1^m(\Theta, \varphi)\, d\Omega}{\int_\Omega \left[\rho_1 \int_0^{r(\Theta, t)} r^2 dr + \rho_2 \int_{r(\Theta, t)}^L r^2 dr\right] d\Omega} = 0.$$

Integrating it over the radial coordinate yields

$$\frac{\frac{L}{4}\int_\Omega \left[\rho_2 + (\rho_1 - \rho_2)\frac{r^4(\Theta, t)}{L^4}\right] Y_1^m(\Theta, \varphi)\, d\Omega}{\frac{1}{3}\int_\Omega \left[\rho_2 + (\rho_1 - \rho_2)\frac{r^3(\Theta, t)}{L^3}\right] d\Omega} = 0.$$

Note that the denominator of this expression is a finite quantity, since (see (5))

$$\int_\Omega d\Omega = 4\pi; \quad \int_\Omega r^3(\Theta, t) d\Omega = 4\pi R^3,$$

and the first integral in the numerator is equal to unity because of the well-known property

$$\int_\Omega Y_1^m(\Theta, t) d\Omega = 0$$

of spherical functions.

Eventually, the condition of center-of-mass immobility for the drop–environment system can be written as

$$\frac{3(\rho_1 - \rho_2)}{16\pi\left[\rho_2 + (\rho_1 - \rho_2)\frac{R^3}{L^3}\right]}\int_\Omega \frac{r^4(\Theta, t)}{L^3} Y_1^m(\Theta, t) d\Omega = 0. \quad (8)$$

It is obvious that, at a sufficiently large linear size L of the environment, equality (8) may hold for any arbitrary function $r(\Theta, t)$ with an accuracy as high as desired.

Thus, the condition of center-of-mass immobility in the problem of drop surface vibration in an environment that occupies a sufficiently large but finite volume is fulfilled automatically. Hence, the translational mode amplitude in expansion (7) should be determined from boundary conditions (2)–(4). Note that, as for the drop in a vacuum, the excitation of the translational mode compensates for the displacement of the center-of-mass of the drop, which is caused by surface vibrational modes [9].

In the absence of the environment ($\rho_2 = 0$), Eq. (8) turns into

$$\frac{3}{16\pi}\int_\Omega \frac{r^4(\Theta, t)}{R^3} Y_1^m(\Theta, t) d\Omega = 0.$$

This condition is not obvious and is usually taken into account in the complete formulation of the problem of drop surface vibrations in a vacuum [4, 7–9].

CONCLUSIONS

Upon solving the problem of the nonlinear vibrations of incompressible ideal liquid drops placed in an immiscible incompressible ideal medium, the condi-

tion of center-of-mass immobility is satisfied automatically. Therefore, the translational mode amplitude should be calculated with a set of hydrodynamic boundary conditions at the interface.

ACKNOWLEDGMENTS

REFERENCES

1. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).

2. A. N. Zharov and S. O. Shiryaeva, Elektron. Obrab. Mater., No. 6, 9 (1999).

3. A. I. Grigor'ev, Zh. Tekh. Fiz. **70** (5), 22 (2000) [Tech. Phys. **45**, 543 (2000)].

4. J. A. Tsamopoulos and R. A. Brown, J. Fluid Mech. **127**, 519 (1983).

5. O. A. Basaran and L. E. Scriven, Phys. Fluids A **1**, 795 (1989).

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

7. S. O. Shiryaeva, Zh. Tekh. Fiz. **71** (2), 27 (2001) [Tech. Phys. **46**, 158 (2001)].

8. S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 163 (2001).

9. S. O. Shiryaeva, Zh. Tekh. Fiz. **72** (4), 15 (2002) [Tech. Phys. **47**, 389 (2002)].

10. S. O. Shiryaeva, Zh. Tekh. Fiz. **73** (2), 19 (2003) [Tech. Phys. **48**, 152 (2003)].

11. T. B. Benjamin and A. T. Ellis, J. Fluid Mech. **212**, 65 (1989).

12. Z. Feng, J. Fluid Mech. **333**, 1 (1997).

13. A. I. Grigor'ev, S. O. Shiryaeva, A. R. Gaibov, *et al.*, Pis'ma Zh. Tekh. Fiz. **27** (22), 7 (2001) [Tech. Phys. Lett. **27**, 934 (2001)].

14. S. O. Shiryaeva, A. I. Grigor'ev, D. F. Belonozhko, and A. S. Golovanov, Pis'ma Zh. Tekh. Fiz. **27** (20), 65 (2001) [Tech. Phys. Lett. **27**, 875 (2001)].

15. M. Kornfeld and L. Suvorov, J. Appl. Phys. **15**, 495 (1944).

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

Translated by K. Shakhlevich

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

Beam–Plasma Mechanism for Anode Plasma Generation in a Low-Pressure Two-Stage Self-Sustained Discharge with a Cold Hollow Cathode

A. I. Stognij* and N. N. Novitskii**

* Zavod Transistor Unitary Enterprise, Integral Research and Production Association, Minsk, Belarus
e-mail: stognij@ifttp.bas-net.by
** Institute of Physics of Solids and Semiconductors, National Academy of Sciences, Minsk, 220072 Belarus
Received December 4, 2002

Abstract—Results are presented from experimental studies of the anode region of a low-pressure two-stage self-sustained discharge with a closed cold hollow cathode. It is shown that applying an external longitudinal magnetic field promotes the generation of a dense anode plasma, whereas the transverse field impedes this generation. It is established that the beam–plasma mechanism for plasma generation plays a dominant role in the anode region of the discharge. The geometry of the electrodes of the gas-discharge chamber is optimized. © 2003 MAIK "Nauka/Interperiodica".

INTRODUCTION

It is known [1–3] that, in a low-pressure glow discharge with a cold hollow cathode, an inhomogeneous dense plasma is produced in the anode region if the following condition is satisfied: $S_a/S_k < 0.01$, where S_a is the anode area and S_k is the area of the inner surface of the hollow cathode. A bright anode plasma glow with regular boundaries is usually observed against the background of the uniform glow of the hollow-cathode positive column. The mechanism for the generation of this dense anode plasma has been investigated over several years. Based on the conventional model of binary collision of fast electrons with neutrals, it is difficult to explain how the cold anode plasma is generated in a volume smaller than 10 cm^3 under typical discharge conditions (the working gas pressure in the anode region is less than 10 Pa, the discharge current is higher than 0.1–1.0 A, and the discharge voltage is 300–600 V). Estimates show that, in this case, the electron mean free path exceeds (or is comparable with) the characteristic dimensions of the anode region, whereas the probability of an ionizing collision for a fast electron in the anode region is significantly less than unity. Moreover, due to a substantial reduction in the relative anode area at low gas pressures, an electric double layer with a potential drop of about several tens of volts is usually formed between the anode and cathode; the field of this layer accelerates the plasma electrons toward the anode [1, 2]. The ion and electron currents through a steady-state double layer separating the anode and cathode plasmas satisfy the well-known relationship: $j_i = (m/M)^{1/2}j_e$, where j_i and j_e are the ion and

electron current densities, respectively, and M and m are the masses of an ion and electron, respectively [4]. It follows from here that, for a typical discharge current of ~10^3 mA, the ion emission current from a self-sustained discharge can hardly exceed 10 mA. However, it is well known that the low-pressure two-stage self-sustained discharge with a closed cold hollow cathode has found applications as an efficient oxygen ion emitter capable of producing ion beams with a current of up to 10–100 mA at discharge currents of 100–1000 mA, in spite of the small anode area and the presence of an ~40-eV electric double layer in the discharge plasma [5, 6]. Experimental investigations show that conditions in the anode region of such a discharge are favorable for collective interactions between the beam of the cathode-plasma electrons accelerated in the electric double layer and the dense anode plasma [2]. However, up to now, there has been no experimental evidence that the beam–plasma interactions play a dominant role in the generation of the cold anode plasma. If the latter hypothesis is true, then, as was shown in [7–9], there is no contradiction between the observed high value of the ion emission current from the discharge and the limits that are imposed by the condition for the currents flowing through the double layer. According to [7–9], the density j_{bi} of the ion emission current from a beam–plasma discharge initiated by injecting an electron beam toward the anode substantially exceeds the density j_i of the ion emission current through a steady-state double layer and is determined by the relationship $j_{bi} \approx \alpha j_i$, where α is the space-charge neutralization factor, which is equal to 5–100, depending on the experimental conditions. In the present paper, conditions under

which the anode region in a low-pressure two-stage self-sustained discharge with a closed cold hollow cathode is formed are compared for three different cases: (i) when an external magnetic field is absent, (ii) when the anode region is in an external longitudinal magnetic field, and (iii) when it is in a transverse magnetic field. It is well known that the longitudinal magnetic field parallel to the direction of electron beam injection increases the intensity of the beam–plasma interaction [10], whereas the transverse magnetic field increases the probability of electron-impact ionization in binary collisions [4]. It is also known that applying a transverse magnetic field to a double layer increases the ion current through it [11].

An analysis of the results presented in this paper (which is a continuation of [2]) points to the dominant role of the beam–plasma ionization mechanism in the formation of the anode plasma in a low-pressure two-stage self-sustained discharge with a cold hollow cathode. To date, conditions under which the beam–plasma discharge is excited and evolves have been investigated only for the cases of a low-voltage (<10 V) beam–plasma discharge at relatively high pressures (>10^2 Pa) in the absence of an external magnetic field [12] and a

higher voltage (>100 V) low-pressure (<10 Pa) beam-plasma discharge both with and without a longitudinal magnetic field [7–10].

EXPERIMENTAL TECHNIQUE

A schematic of the experimental discharge chamber is shown in Fig. 1. The discharge was excited between the inner surface of the closed hollow cathode (electrodes *1–3*) and plane anode (*4*) through a contraction hole (*5*). The height of the walls of the contraction hole was 3 mm, and the hole diameter *d* varied from 4 to 14 mm with the help of accessory inserts. The height *h* of the gap between the anode (*4*) and the hollow-cathode wall (*1*) varied from 1 to 8 mm. The electrodes were made of nonmagnetic stainless steel. A longitudinal magnetic field $B_{\parallel}$ of up to ~15 mT in the anode–cathode gap was created by two standard axially magnetized ferrite–barium rings (*6, 7*) mounted coaxially in aluminum shields at the anode (*4*) and the inner surface of the hollow-cathode wall (*1*), respectively. A transverse magnetic field B_r of up to 20 mT in the anode–cathode gap was created by two longitudinally magnetized ferrite–barium rods (*8, 9*) mounted at the outer surface of the hollow-cathode wall (*1*) symmetrically about the contraction hole (*5*). As in the prototype device [2], the cylindrical wall (*2*) of the hollow cathode was 120 mm in diameter and 80 mm in height and had windows shielded with a molybdenum grid. The discharge chamber was pumped at a rate of ~2 m²/s to a pressure of ~10^{-4} Pa through 520 3-mm-diameter holes in the bottom of the cathode (*3*). Oxygen was fed into the discharge chamber through a peripheral hole in the support of the anode (*4*). At an oxygen flow rate *Q* of 1.1×10^{-2} Pa m³/s, the pressure in the anode–cathode gap was less than 2 Pa, whereas in the hollow cathode it was less than 0.1 Pa. The discharge in the anode–cathode gap was photographed through a sealed quartz insulator (*10*). The discharge chamber was powered from a stabilized rectifier with an output current of 1 A and output voltage of 1 kV. The anode circuit included a 100-Ω ballast resistor. The electrodes of the hollow cathode were grounded.
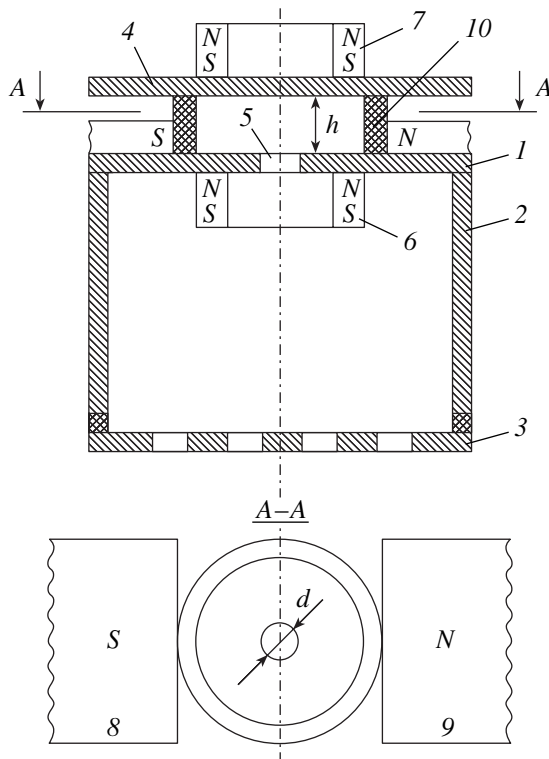

Fig. 1. Experimental discharge chamber: (*1*) upper cathode wall, (*2*) cylindrical wall of the hollow cathode, (*3*) lower cathode wall, (*4*) plane anode, (*5*) contraction hole, (*6, 7*) annular magnets, (*8, 9*) magnet rods, and (*10*) quartz insulator.

EXPERIMENTAL RESULTS AND DISCUSSION

Figure 2 shows a set of photographs of the discharge glow in the anode region under different experimental conditions. The anode plasma in the absence of a magnetic field (Fig. 2a) and at characteristic values of the discharge parameters (the current $I_d = 0.3$ A, the discharge voltage $U_d = 360$ V, the oxygen flow rate $Q = 1.2 \times 10^{-2}$ Pa m³/s, the height of the anode–cathode gap $h = 8$ mm, and the diameter of the contraction hole $d = 4$ mm) resembles a typical beam–plasma discharge in the absence of a magnetic field, because it contains all the main components of the latter (Fig. 2b) [13]. By analogy with [13], the following regions can be distin-

guished in the anode plasma: a cylindrical plasma column C produced by a nearly parallel electron beam from the cathode plasma; a meniscus region E; and a scattering region F, in which the beam electrons are deflected in the radial direction. As the height of the anode–cathode gap decreases, the discharge voltage U_d somewhat decreases (the other discharge parameters remaining the same); the diameter of the scattering region F increases; and the cylindrical plasma column C is displaced inside the hollow cathode, where the pressure is lower. This is illustrated by Fig. 2c, which corresponds to $I_d = 0.3$ A, $U_d = 345$ V, $h = 4$ mm, and $d = 4$ mm. As the discharge current was increased by supplying a higher power from the external source, the discharge voltage increased and the scattering region F grew in size (see Fig. 2d, in which $U_d = 350$ V and $I_d = 0.4$ A). Applying the transverse magnetic field B_r deteriorated the discharge conditions. In this case, the intensity of the anode plasma glow decreased, the diameter of the plasma column increased, and its boundaries became irregular (see Fig. 2e, which corresponds to $I_d = 0.3$ A, $U_d = 375$ V, $Q = 1.2 \times 10^{-2}$ Pa m^3/s, $h = 8$ mm, and $d = 6$ mm). As the transverse magnetic field was increased to $B_r = 20$ mT, the discharge conditions even more degraded and the intensity of the anode plasma glow further decreased (see Fig. 2f, in which $I_d = 0.3$ A and $U_d = 390$ V). At the same time, the discharge became unstable and the generation of intense broadband noise was observed. At an initial discharge current (without an external magnetic field) of $I_d \leq 0.2$ A and oxygen flow rate of $Q \leq 1.2 \times 10^{-2}$ Pa m^3/s, a magnetic field of $B_r = 20$ mT was sufficient to suppress the discharge. An opposite situation took place when a longitudinal magnetic field was imposed. In this case, even with a longitudinal magnetic field in the anode–cathode gap of $B_\parallel \geq 10$ mT, we observed an intense uniform anode plasma glow with well-defined side boundaries throughout the entire anode–cathode gap (see Fig. 2g, which corresponds to $I_d = 0.3$ A, $U_d = 350$ V, $Q = 1.2 \times 10^{-2}$ Pa m^3/s, $h = 8$ mm, and $d = 4$ mm). As the discharge current was increased to $I_d = 0.6$ A, the discharge voltage insignificantly increased (to $U_d = 365$ V) and the anode plasma glow became more intense and uniform (Fig. 2h).

Figure 3 shows the current–voltage characteristics $U_d(I_d)$ of the discharge (curves 1–3) and the dependences of the current I_k flowing through the lower wall (3) of the hollow cathode on the discharge current I_d (curves 4–6, characterizing the efficiency of ion emission from the discharge) for the three types of the discharge conditions considered above: without an external magnetic field in the anode region (curves 1, 4), with a transverse magnetic field of $B_r = 15$ mT (curves 2, 5), and with a longitudinal magnetic field of $B_\parallel = 15$ mT (curves 3, 6). The other parameters are $Q = 1.4 \times 10^{-2}$ Pa m^3/s, $d = 6$ mm, and $h = 4$ mm. It can be seen that, in the discharge current range under consideration
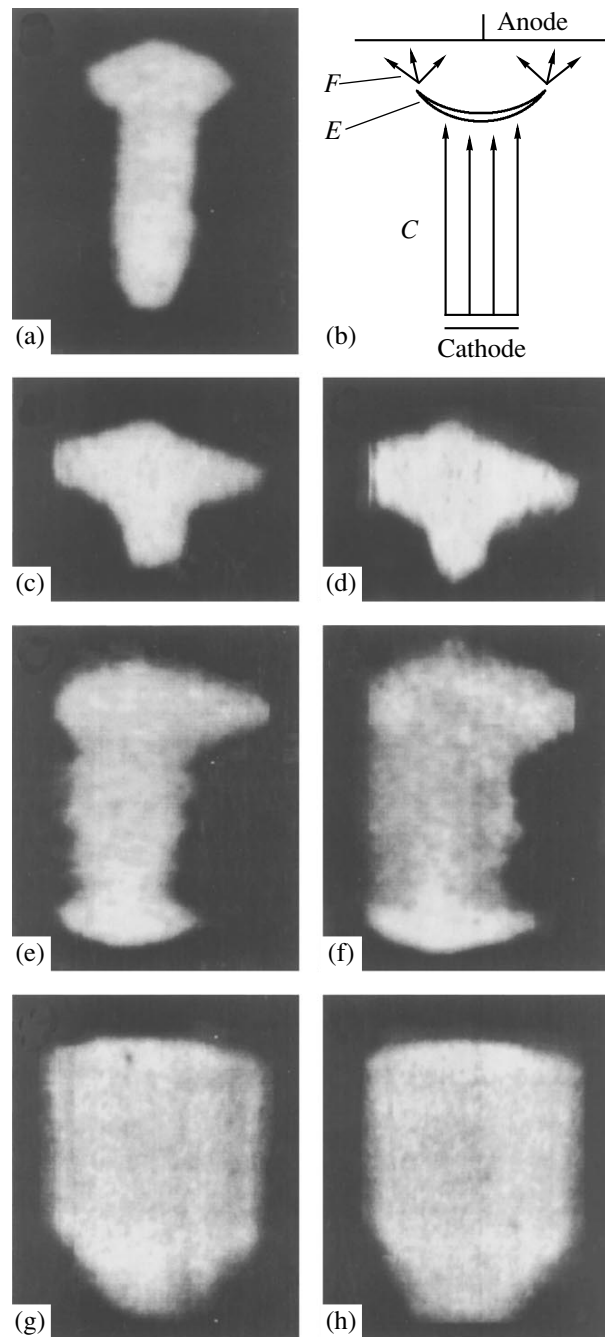

Fig. 2. Photographs of the anode region of the discharge (a, c, d) in the absence of a magnetic field, at $B_r =$ (e) 10 and (f) 20 mT, and (g, h) at $B_\parallel = 15$ mT. Plot (b) shows a schematic of the beam–plasma interaction region [13]: C is the plasma column, E is the meniscus region, and F is the scattering region.

(from 0.2 to 0.6 A) and in the absence of an external magnetic field, the discharge voltage U_d increases with the current I_d and the dependence $I_k(I_d)$ saturates at $I_d > 0.3$ A. In the case with a transverse magnetic field, the slope of the curve $U_d(I_d)$ increases sharply at discharge currents of $I_d > 0.3$ A and the current I_k at these values
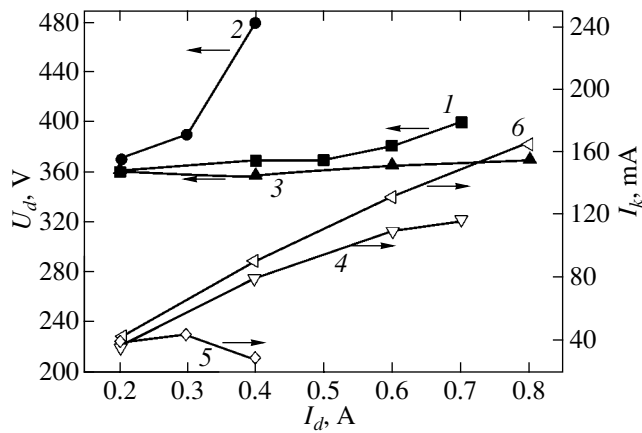
Fig. 3. (*1–3*) Current–voltage characteristics of the discharge and (*4–6*) the current flowing through the lower wall of the hollow cathode as a function of the discharge current (*1, 4*) in the absence of a magnetic field, (*2, 5*) at $B_r = 15$ mT, and (*3, 6*) at $B_\parallel = 15$ mT.

of the discharge current starts to gradually decrease. In the case with a longitudinal magnetic field, the dependence $U_d(I_d)$ is presented by a horizontal line, whereas the curve $I_k(I_d)$ monotonically increases.

The data presented in Figs. 2 and 3 testify to the beam–plasma mechanism for anode plasma generation. Indeed, in the absence of an external magnetic field, the energy acquired by the electrons accelerated in the electric field of the double layer and the current density in the electron beam arriving from the cathode plasma can be insufficient to generate a dense nonequilibrium plasma in the anode region in spite of an increase in the power supplied from the external source. The reason may be, e.g., the loss of electrons on the anode and the loss of ions on the cathode wall near the contraction hole. The fact that the fast-electron scattering region is located near the anode, as well as the large cross size of this region (see Figs. 2c, 2d), leads to a decrease in the probability of anode plasma ions falling into the contraction hole, which hinders the increase in the ion current flowing through the electric double layer as the discharge current increases. That is why the dependence $U_d(I_d)$ is monotonically increasing and the curve $I_k(I_d)$ is saturated. The transverse magnetic field magnetizes the electrons moving toward the anode; confines fast electrons near the contraction hole, where the working gas pressure is maximum; and randomizes the motion of fast electrons toward the anode. As a result, the conditions for beam–plasma interaction in the anode region deteriorate, whereas the conditions for ionization due to binary collisions improve [4]. In addition, the transverse magnetic field promotes the ion current to flow through the double layer, because the space charge of the double layer is partially neutralized by cathode-stage electrons oscillating along the magnetic field lines [11]. Nevertheless, as the working pressure decreases and/or the magnetic field increases, the dis-

charge conditions deteriorate and the discharge becomes located in the higher pressure region (near the contraction hole) and even takes the form of a low-current diode discharge in a transverse magnetic field (Figs. 2e, 2f). As the discharge current increases, the current–voltage characteristic $U_d(I_d)$ becomes nearly vertical. The decreasing dependence $I_k(I_d)$ reflects a decrease in the probability of the discharge ions reaching the lower wall of the hollow cathode (Fig. 3, curves *3, 6*). In contrast, the longitudinal magnetic field impedes the transverse loss of the fast cathode-plasma electrons in the anode plasma region. As a result, favorable conditions for collective interactions in the anode region are created [10] and a dense homogeneous anode plasma is generated (Figs. 2g, 2h). At the same time, the discharge conditions are improved: the current–voltage characteristic becomes almost horizontal, and the value of the ion current to the lower wall of the hollow cathode increases with increasing discharge current (Fig. 3, curves *2, 5*).

Combining the results from theoretical models of beam–plasma interactions and analyzing actual experimental conditions in beam–plasma discharge devices, a number of semiempirical relations allowing one to optimize the geometry of the discharge chamber was obtained in [5, 8, 10]. Thus, for the optimum geometry of a waveguide-type discharge chamber in an external longitudinal magnetic field, the ratio of the length L to the radius R of the beam–plasma interaction region (it is assumed that $L \gg R$) should satisfy the condition [8]

$$\frac{L}{R} = \varepsilon_{nm} V_b^{3/2} (e/2m_e)^{1/2}/I_b.$$

Here, ε_{nm} is the nth zero of the Bessel function of mth order, $J_m(x)$; e and m are the electron charge and mass, respectively; V_b is the potential of the electric field accelerating the beam electrons (in our case, the field of the double layer); and I_b is the ion beam current. It follows from this formula that, for $L \approx$ const and $V_b \approx$ const, the radius R of the interaction region should be increased as I_b increases in order to maintain the optimum conditions for the beam–plasma interaction. Extending this conclusion to the case of the anode region of a two-stage self-sustained discharge (in spite of the fact that the length of this region is only several-fold greater than its radius), we find that more favorable conditions for a high-current discharge can be achieved by increasing the diameter of the contraction hole as the discharge current increases. The same conclusion can be drawn from Figs. 2e and 2g, in which the size of the dense homogeneous anode plasma is nearly twice as great as the diameter of the contraction hole.

These considerations stimulated the optimization of the geometry of the anode region of the discharge chamber by optimizing the geometry of the contraction hole. For this purpose, several inserts with contraction holes of different geometry were used. Figure 4 shows the change in the geometry of the contraction hole after
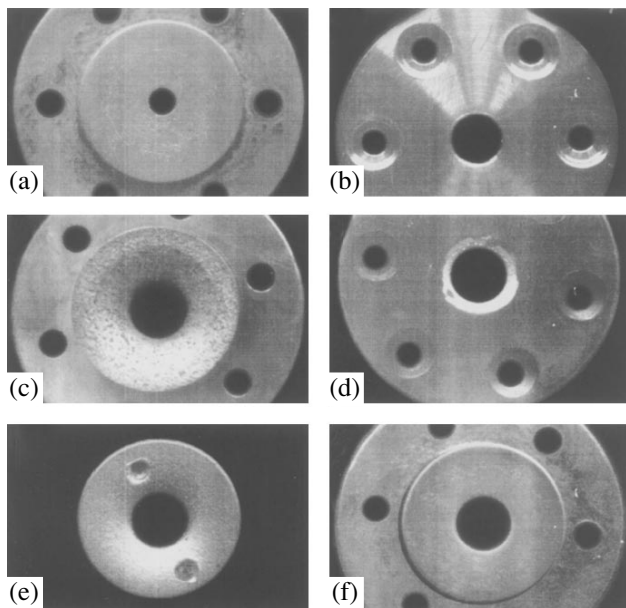
Fig. 4. Change in the geometry of the contraction hole due to the wall sputtering under different conditions in the anode region of the discharge: (a, b) in the absence of a magnetic field, (c) at $B_r = 15$ mT, and (d–f) at $B_{\parallel} = 15$ mT.

long-term discharge operation. In all cases, the discharge current and the oxygen flow rate were maintained at a constant level: $I_d = 0.4$ A and $Q = 1.2 \times 10^{-4}$ Pa m^3/s. The length of the anode–cathode gap was $h = 4$ mm. These experiments have shown that, in the absence of an external magnetic field, the hole diameter increased due to the wall sputtering from 6 mm (the initial contraction hole is shown in Fig. 4a) to ~12 mm over a time interval of 200 h. As a result, a conical crater with a base on the side of the hollow cathode was formed (Fig. 4b). As the diameter of the contraction hole increased further, the discharge conditions deteriorated and the discharge evolved into a high-voltage discharge. When a transverse magnetic filed of $B_r = 15$ mT was applied, the wall was sputtered more intensely and, after 85-h operation, the discharge became unstable, the diameter of the contraction hole increased from 6 to 13 mm, and its wall became eroded from both the anode and cathode sides (the resulting state of the contraction hole is shown in Fig. 4c). A different picture was observed in the case of a longitudinal magnetic field. For a longitudinal magnetic field of $B_{\parallel} = 15$ mT, the optimum diameter of the contraction hole was $d = 12$ mm (the initial contraction hole is shown in Fig. 4d). The diameter of the contraction hole remained unchanged even after 200-h discharge operation; however, the wall edges turned out to be noticeably eroded on the anode side (Fig. 4e), whereas they were only slightly eroded on the cathode side (Fig. 4f). With lower initial diameters of the contraction hole and in the presence of a longitudinal magnetic field, the discharge conditions somewhat improved (the other discharge

parameters being the same, the discharge voltage decreased by no more than 10 V); however, the wall of the contraction hole on the anode side was sputtered more intensely. With hole diameters larger than $d = 12$ mm, the discharge conditions deteriorated. The coincidence of the values of the hole diameter ($d \geq 12$ mm) at which a considerable deterioration of the discharge conditions was observed in all the above three cases (regardless of the orientation of the external magnetic field) may be related to the necessity of satisfying the condition for the existence of a double layer, $S_a/S_k < 0.01$, or in less explicit form, the necessity of adjusting the surface area of the double layer to the diameter of the contraction hole. It is obvious that this condition fails to be satisfied as the diameter of the contraction hole increases. The results of optimizing the geometry of the anode region of the discharge were used to design oxygen-ion sources producing 300-mm ion beams with a uniform current density over the beam cross section [6].

CONCLUSION

The experimental results presented in this paper allow one to consider the anode region of a low-pressure two-stage self-sustained discharge with a cold hollow cathode as a low-voltage non-self-sustained beam–plasma discharge in a longitudinal magnetic field. The discharge is initiated by the beam of the cathode-plasma electrons, which are accelerated toward the anode by the field of an electric double layer located on the cathode side of the contraction hole in the hollow cathode. This conclusion is confirmed by the results of optimizing the geometry of the anode region of the discharge.

ACKNOWLEDGMENTS

REFERENCES

1. A. S. Metel', Zh. Tekh. Fiz. **54**, 241 (1984) [Sov. Phys. Tech. Phys. **29**, 141 (1984)].

2. A. I. Stognij, V. A. Nikitinskiĭ, and B. I. Zhuravlev, Zh. Tekh. Fiz. **58**, 993 (1988) [Sov. Phys. Tech. Phys. **33**, 604 (1988)].

3. O. A. Nerushev, S. A. Novopashin, V. V. Radchenko, and G. I. Sukhinin, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 679 (1997) [JETP Lett. **66**, 711 (1997)].

4. V. L. Granovskiĭ, *Electric Current in Gas: Steady-State Current* (Nauka, Moscow, 1971).

5. A. I. Stognij, V. A. Nikitinskiĭ, and B. I. Zhuravlev, USSR Inventor's Certificate No. 1561744 (1991), Byull. Izobret., No. 33 (1991), p. 215.

6. A. I. Stognij and S. V. Koryakin, Prib. Tekh. Éksp., No. 6, 64 (2000).

7. R. A. Demirkhanov, Yu. V. Kursanov, and L. P. Skripal', Zh. Tekh. Fiz. **40**, 1351 (1970) [Sov. Phys. Tech. Phys. **15**, 1047 (1970)].

8. M. G. Rosing and J. P. Conrad, J. Appl. Phys. **57**, 816 (1985).

9. *The Physics and Technology of Ion Sources*, Ed. by I. G. Brown (Wiley, New York, 1989; Mir, Moscow, 1988).

10. A. A. Ivanov, Itogi Nauki Tekh., Ser. Fiz. Plazmy **3**, 176 (1982).

11. V. A. Nikitinskiĭ, B. I. Zhuravlev, and A. T. Gaponenko, Zh. Tekh. Fiz. **55**, 1637 (1985) [Sov. Phys. Tech. Phys. **30**, 951 (1985)].

12. F. G. Baksht, A. A. Bogdanov, V. B. Kaplan, *et al.*, Fiz. Plazmy **7**, 547 (1981) [Sov. J. Plasma Phys. **7**, 296 (1981)].

13. V. P. Kovalenko, Usp. Fiz. Nauk **139**, 223 (1983) [Sov. Phys. Usp. **26**, 116 (1983)].

Translated by N. Larionova

GAS DISCHARGES, PLASMA

Scaling Laws for the Spatial Distributions of the Plasma Parameters in the Positive Column of a DC Oxygen Discharge

E. A. Bogdanov*, A. A. Kudryavtsev*, L. D. Tsendin**, R. R. Arslanbekov***, V. I. Kolobov***, and V. V. Kudryavtsev****

* St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 199034 Russia
e-mail: akud@ak2138.spb.edu
** St. Petersburg State Technical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia
*** CFD Research Corporation, 215 Wynn Drive, Huntsville, AL, USA
**** CFD-Canada, 45 English Ivyway, Toronto, Canada
Received March 4, 2003

Abstract—Comprehensive self-consistent simulations of the positive column plasma of a dc oxygen discharge are performed with the help of commercial CFDRC software (http://www.cfdrc.com/~cfdplasma), which enables one to carry out computations in an arbitrary 3D geometry using fluid equations for heavy components and a kinetic equation for electrons. The main scaling laws for the spatial distributions of charged particles are determined. These scaling laws are found to be quite different in the parameter ranges that are dominated by different physical processes. At low pressures, both the electrons and negative ions in the inner discharge region obey a Boltzmann distribution; as a result, a flat profile of the electron density and a parabolic profile of the ion density are established there. In the ion balance, transport processes prevail, so that ion heating in an electric field dramatically affects the spatial distribution of the charged particles. At elevated pressures, the volume processes prevail in the balance of negative ions and the profiles of the charged particle densities in the inner region turn out to be similar to each other. © *2003 MAIK "Nauka/Interperiodica".*

Interest in discharges in electronegative gases stems from their wide use in modern plasma technologies [1]. In order to predict the possible parameter distributions and their dependence on the external conditions, considerable attention is paid to elucidating the relations between the main plasma parameters. Various aspects of this problem as applied to electronegative gases were considered by many research groups (see, e.g., [1–26]). It was found that, in the presence of negative ions, the processes of spatial transport, which determine the density profiles and other plasma parameters, possess a number of specific features [26]. Knowledge of the spatial distributions of charged particles is of crucial importance for understanding and optimizing the operation of various devices and technologies, such as ion sources and facilities for the plasma treatment of materials. Early attempts to reduce the problem to a set of ambipolar diffusion coefficients by using simplified models [2–9] were contradictory and there were no criteria for their applicability. In [10–12], it was shown that a specific feature of an electronegative-gas plasma is that it stratifies into regions with different ion compositions. In the outer region (shell) of such a plasma, negative ions are practically absent (Figs. 1–3), because they are drawn by the electric field into the plasma interior. Although the thickness of this shell is usually small, its presence is of fundamental importance because it confines the negative ions inside the plasma volume. As a result, the flux of negative ions to the wall

is practically absent (in contrast to those of electrons and ions). In such a situation, the only means to extract negative ions from the discharge is to apply an accelerating voltage U to the wall (or an extracting electrode). The magnitude of this voltage should be large enough for the space charge layer produced at the plasma boundary to extend to the inner region containing negative ions. The thicker the shell, the higher voltage ($U \sim$
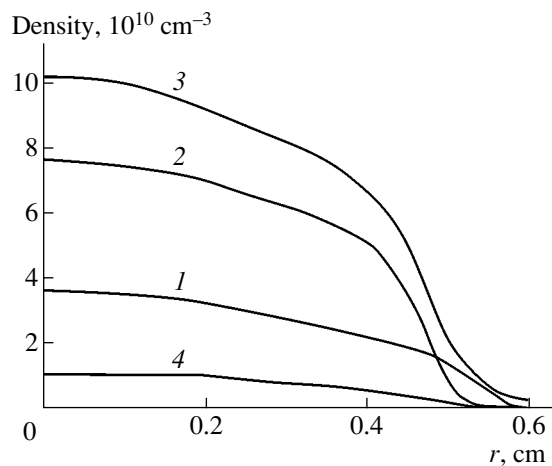

Fig. 1. Profiles of the charged particle densities for $p = 1$ torr and $I = 50$ mA: (*1*) n_e, (*2*) n_n, (*3*) n_p, and (*4*) $n[O^+]$.
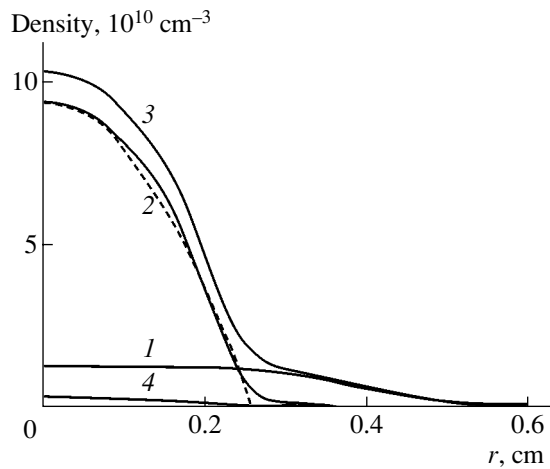
Density, 10^{10} cm^{-3}

Fig. 2. Same as in Fig. 1 for $p = 0.15$ torr without allowance for ion heating: (*1*) n_e, (*2*) n_n, (*3*) n_p, and (*4*) $n[\mathrm{O}^+]$. The dashed curve shows parabolic distribution (18).
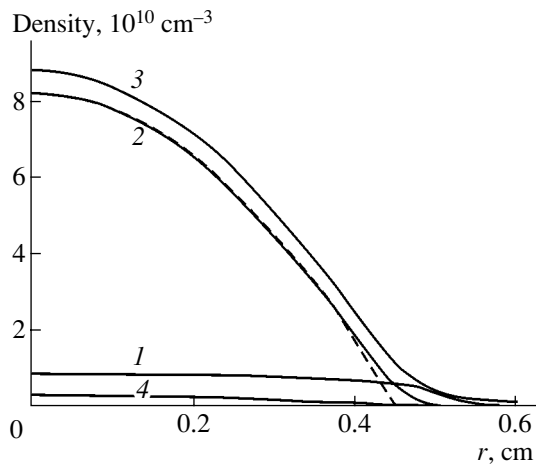
Density, 10^{10} cm^{-3}

Fig. 3. Same as in Fig. 2, but with allowance for ion heating.

$L_{\mathrm{sh}}^{2/3}$) that must be applied to enable the flux of negative ions to the wall.

Besides its practical importance, oxygen plasma is also an important test object [16]. The peculiar features of the spatial distributions of the oxygen plasma parameters have been the subject of a heated discussion. Thus, in [16–18], it was pointed out that using a Boltzmann distribution from [4] for not only electrons but also negative ions (as was done in [13–15]) is unjustified. It is alternatively asserted in [16–18] that the densities of charged particles in such a plasma should be proportional to each other [10–12]. In [24], it was shown that, depending on conditions, both types of distribution can occur. For example, we observed a transition from one type of the above profiles to another as the pressure was reduced (see below).

To verify the functional relations between the plasma parameters obtained with the help of simplified

models, they should be compared with the results of full-scale numerical simulations. Such simulations should be based on self-consistent models that take into account spatial transport processes and volume plasmochemical reactions. Such an attempt was made by us in [27], in which we compared the results of kinetic and fluid simulations of the positive column plasma of a dc oxygen discharge by using commercial CFDRC software [28]. A two-temperature (2T) fluid model was proposed, which allowed us to incorporate kinetic effects in the conventional fluid model in the simplest way.

Here, we continue the study of [27]; specifically, we investigate the features of the spatial distributions of the plasma parameters in the positive column of a dc discharge in a 12-mm-diameter glass tube at pressures of 0.05–3 torr and discharge currents of 5–200 mA. These conditions correspond to those in [21, 22], in which, in our opinion, one of the most detailed experimental and theoretical studies of the positive column of a dc oxygen discharge were reported.

The discharge was simulated by using a commercial software developed at the CFD Research Corporation (Huntsville, AL, USA) [28]. A detailed self-consistent model of the discharge plasma, numerical iteration scheme, and technique for solving the set of equations are described in [28]. The density and mean energy of the electron component can be obtained by solving either fluid balance equations or the kinetic equation for the electron distribution function (EDF). The self-consistent electric field is found from Poisson's equation. Heavy particles are described in the fluid model. Both the analytic results and the published data show that, in the parameter range under study, the neutral gas is heated to no higher than 50–150 K. Such an increase in the gas temperature T results merely in a decrease in the gas density. Since this is of minor importance for our problem, the gas temperature was assumed to be equal to room temperature and constant over the discharge cross section. On the other hand, the ion temperature can increase significantly, particularly at low pressures [20, 21]. As the pressure decreases, the reduced electric field E/p increases, so that the directed velocity acquired by the ions in this field can become higher than the random (thermal) velocity [29]. The coefficient of ion diffusion also increases. This can dramatically change the ion density profiles [20]. Model calculations with allowance for ion heating in a longitudinal electric field show that the outer region occupied by the electron–ion plasma shrinks and can even completely disappear [20, 21]. Ion heating also leads to a decrease in the detachment rate constant and, consequently, to an increase in the relative density of negative ions n_n/n_e (the degree of electronegativity) [21].

Here, we do not present the list of the volume plasmochemical reactions involved because it is the same as in [27]. Note only that we solved the balance equations for the vibrationally excited states $\mathrm{O}_2(v)$ ($v = 0$, 1)

of an oxygen molecule and the electronically-excited states of an oxygen molecule ($O_2(X^3\Sigma_g^-)$, $O_2(a^1\Delta)$, $O_2(b^1\Sigma)$, and $O_2(Ry)$); oxygen atom ($O(^3P)$, $O(^1S)$, and $O(^1D)$); ozone molecule O_3; and O^+, O_2^+, O_4^+, O^-, O_2^-, and O_3^- ions with allowance for 160 plasmochemical reactions between them.

Typical ion density profiles computed for the gas pressure $p = 1$ torr are shown in Fig. 1, and those computed for $p = 0.15$ torr without and with allowance for ion heating in the longitudinal electric field are shown in Figs. 2 and 3, respectively. Ion heating [20–22] was calculated by the formulas for the effective transverse ion temperature [29]

$$T_i^\perp = T + \frac{(M_i + M)Mw^2}{3(2M + M_i)}, \tag{1}$$

where M and M_i are the masses of a molecule and an ion, respectively, and w is the ion drift velocity in the longitudinal electric field E_z.

For example, at $p = 1$ torr, the transverse ion temperature is $T_i^\perp \approx 760$ K, whereas at $p = 0.15$ torr, it is $T_i^\perp \approx 5200$ K. For oxygen, an order of magnitude of the ion temperature as a function of the parameter $p\Lambda$ is presented, e.g., in [21, Fig. 5].

It can be seen from Figs. 1–3 that the spatial distribution of the charged particle densities is highly nonuniform over the discharge cross section. Almost all of the negative ions reside in the inner ion–ion plasma region (which will be marked by subscript 0). The radius of this region is $r = r_0$. The outer electron–ion plasma region ($r_0 < r < R$) (subscript 1) consists of electrons and positive ions, whereas the negative ions are practically absent there. A comparison of the profiles presented in Figs. 2 and 3 show that taking into account ion heating (which increases the ion diffusion coefficient) dramatically changes the shell thickness. For this reason, when analyzing the spatial profiles of the charged particle densities in electronegative gases, one of the central problems is the problem of the ion temperature [20, 21].

The main positive ion is O_2^+ and the main negative ion is O^-. The densities of all other ions are small compared with the densities of these ions. Hence, for the sake of qualitative analysis, it is sufficient to consider a plasma consisting of only electrons, positive ions, and negative ions (subscripts e, p, and n, respectively).

To explain the dependences observed and predict how they are affected by the external conditions, we consider, as in [1–26], the conventional set of drift–diffusion equations

$$-D_p\nabla(\nabla n_p + kn_p\nabla n_e/n_e) = \nu_i n_e - K_r n_n n_p, \tag{2}$$

$$-D_n\nabla(\nabla n_n - kn_n\nabla n_e/n_e) = \nu_a n_e - \nu_d n_n - K_r n_n n_p, \tag{3}$$

$$n_p = n_n + n_e. \tag{4}$$

with a Boltzmann distribution for the electrons: $E = -T_e\nabla n_e/n_e$). Here, ν_i, ν_a, and ν_d are the ionization, attachment, and detachment frequencies, respectively; K_r is the rate constant for ion–ion recombination; and $k = T_e/T_i$ is the electron-to-ion temperature ratio. The boundary conditions for the set of Eqs. (2) and (3) are [1, 26]

$$\nabla n_n = \nabla n_p = 0 \quad \text{at} \quad r = 0,$$
$$n_n = n_p = \nabla n_n = 0 \quad \text{at} \quad r = R. \tag{5}$$

Since the flux of negative ions to the wall is zero, we find from Eq. (3) that the densities averaged over the cross section ($\bar{n}$) satisfy the relationship [1, 26]

$$\nu_a\bar{n}_e = \nu_d\bar{n}_n + K_r\bar{n}_n\bar{n}_p. \tag{6}$$

In further analysis, we will mainly follow [24–27]. We divide Eqs. (2) and (3) by the corresponding diffusion coefficients and sum them up. As a result, we arrive at the equation [10, 26]

$$-2\Delta n_n/k - \Delta n_e = n_e/l_e^2 - 2n_n/kl_n^2, \tag{7}$$

which is of fundamental importance for analyzing the solution to the set of Eqs. (2) and (3). Equation (7) contains two characteristic space scales, l_e and l_n, which are defined by

$$1/l_e^2 = 1/l_{ion}^2 + 1/l_a^2 = \nu_i/D_{ap} + \nu_a/D_{an}$$
$$= \tau_{ap}\nu_i/\Lambda^2 + \tau_{an}\nu_a/\Lambda^2, \tag{8}$$

$$1/l_n^2 = 1/l_{nd}^2 + 1/l_{nr}^2 = \nu_d/2D_n + n_pK_r/D_{np}$$
$$= \tau_n\nu_d/\Lambda^2 + \tau_{np}K_r n_p/\Lambda^2, \tag{9}$$

where $D_{an, ap} = D_{n, p}(k + 1)$ and $D_{np} = 2D_nD_p/(D_n + D_p)$ are the coefficients of electron–ion and ion–ion ambipolar diffusion, respectively; $\tau_j = \Lambda^2/D_j$ are the corresponding characteristic times; and Λ is the diffusion length, which, in the case of cylindrical geometry, is equal to $\Lambda = R/2.4$. Figure 4 shows, as an example, the calculated lengths (8) and (9) versus the parameter $p\Lambda$ for oxygen.

Since the ambipolar electric field draw negative ions into the plasma, their density in the outer region ($r_0 \leq r \leq R$) is low, $n_n(r) \approx 0$; hence, we have $n_e(r) \approx n_p(r)$ in this region. Neglecting the terms with n_n, we can write Eq. (7) in the form

$$-\Delta n_e = n_e/l_e^2.$$

Taking into account the spread caused by ion diffusion, we find that the thickness of the outer region satisfies the condition $R - r_0 \leq l_e$; i.e., l_e determines the maximum thickness of the shell. Under our conditions, this thickness is small compared with the tube radius R (and, hence, with the characteristic diffusion length $\Lambda =$
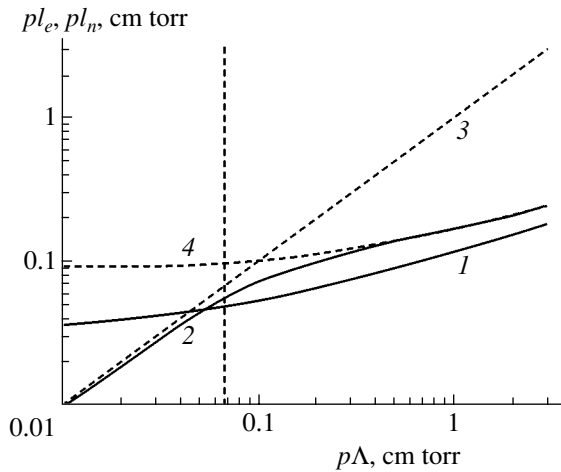
Fig. 4. Characteristic lengths l_n (*1*) and l_e (*2*) vs. parameter $p\Lambda$ for an oxygen discharge. The asymptotes for l_e, $l_e \approx \Lambda$ (*3*) and $l_e \approx l_n \sqrt{T_e/2T_i}$ (*4*) are also shown.

Fig. 5. Contributions of spatial transport and volume processes to the negative ion balance for $p = 1$ torr and $I = 50$ mA. Curve *1* shows the flux of negative ions (with a minus sign), curve *2* shows the total production of ions, and curve *3* shows the ion loss.

$R/2.4$). Therefore, the outer region *1* can be treated in plane geometry. Then, for the plasma density profile in the region $r_0 \le r \le R$, we can use the solution [10, 24–26]

$$n_e(r) = n_e(r_0)\sin(\pi(R-r)/2l_e)/\sin(\pi(R-r_0)/2l_e).\,(10)$$

The density profiles in the inner region *0* depend substantially on the ratio between R and l_n (see Eq. (9)), i.e., between the radius and the distance a negative ion covers due to diffusion during its lifetime with respect to volume processes [24–26]. At $\tau_{an}\nu_a > 1$, length l_e (8) is small ($l_e < \Lambda$), and, under typical discharge conditions ($\bar{n}_n/\bar{n}_e < k \approx 100$), length l_n turns out to be even smaller ($l_n < l_e$); hence, ion diffusion can be ignored [10–12]. When the opposite inequality is satisfied ($\tau_{an}\nu_a < 1$), the electron–ion plasma occupies almost the entire cross section of the tube, whereas length l_n can be either longer or shorter than the radius of the inner ion–ion region. Hence, to obtain functional dependences in the inner region, it is reasonable to consider two limiting regimes with large and small values of the parameter $\tau_{an}\nu_a$. Since this parameter is quadratic in pressure, the boundary between these regimes ($\tau_{an}\nu_a = 1$) can be determined with sufficient accuracy. For oxygen, the boundary value of this parameter corresponds to $p\Lambda \approx 0.07$ cm torr (the dashed vertical line in Fig. 4), so that $\tau_{an}\nu_a > 1$ at $p\Lambda > 0.07$ cm torr and vice versa. Consequently, length l_e (8) has two asymptotes: $l_e \approx \Lambda$ at low pressures, $p\Lambda < 0.07$ cm torr, and $l_e \approx l_n\sqrt{T_e/2T_i}$ in the opposite case (Fig. 4).

At high attachment frequencies ($\tau_{an}\nu_a > 1$), characteristic lengths (8) and (9), as was mentioned above, are both small ($l_n < l_e < \Lambda$) (Fig. 4). Since $l_n < l_e$, we can neglect ion diffusion in Eqs. (2) and (3) (as was done in
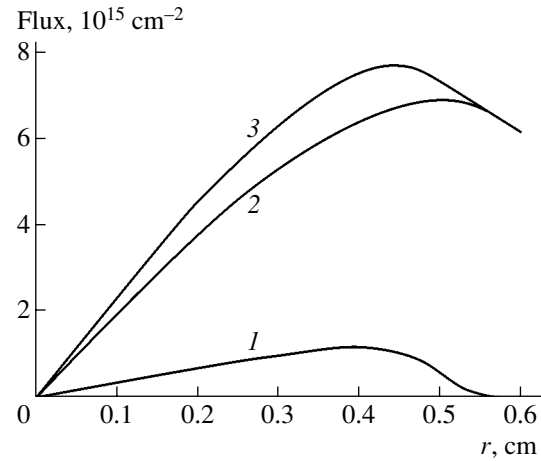
[10–12]) and assume that the shell thickness is $R - r_0 \approx l_e$ (i.e., the denominator in Eq. (10) is equal to unity). In the case at hand, in balance equation (3) for negative ions, their transport is insignificant as compared to volume processes (see Fig. 5), so that the negative ion flux is almost completely determined by the drift component. Hence, at $n_p \approx n_n > n_e$, the fluxes of positive and negative ions in the inner region are almost the same in magnitude, but opposite in sign; i.e., we have [10, 26]

$$\Gamma_n/b_n \approx k n_p \nabla n_e/n_e \approx k n_n \nabla n_e/n_e \approx -\Gamma_p/b_p. \quad (11a)$$

For this reason, in Eq. (7), in which these fluxes are summed up, they almost completely cancel each other in the inner ion–ion region. In other words, at $r < r_0$, the terms on the left-hand side of Eq. (7) (which are responsible for spatial transport) are small compared to the terms on the right-hand side (which are responsible for volume processes). Hence, the local balance of the volume plasmochemical processes resulting in the production and loss of ions, $n_e/l_e^2 = 2n_n/kl_n^2$, holds with a high accuracy. At $\tau_{an}\nu_a > 1$, the following important relation can be deduced from this equality [10, 24–26]:

$$
\begin{aligned}
(\nu_i/D_p + \nu_a/D_n)n_e &= \nu_d n_n/D_n \\
&+ K_r n_n(n_n + n_e)/(1/D_p + 1/D_n),
\end{aligned}
\quad (12a)
$$

which allows one to obtain the relationships between the plasma parameters in the central region $r < r_0$.

The relationships between the densities of charged particles depend on the mechanism responsible for the loss of negative ions, i.e., on the relationship between the terms on the right-hand side of Eq. (12a). At $\tau_{an}\nu_a > 1$, the loss of negative ions in an oxygen plasma is governed by detachment processes (the detachment regime with $\nu_d > n_p K_r$) and their recombination can be neglected. Then, it follows from Eq. (12a) that the pro-

files of the electron and negative ion densities are similar to each other,

$$\nabla n_e/n_e = \nabla n_n/n_n, \quad n_e(x)/n_n(x) = \text{const.} \quad (12b)$$

This condition was first proposed in [2] and then was justified in [10–12], assuming that ion diffusion can be neglected as compared to ion drift. It follows from the above analysis that Eq. (12b) is valid only at $\tau_{an}\nu_a > 1$; hence, extrapolating it to the low pressure range [16–18] is incorrect. The validity of Eq. (12b) for oxygen is illustrated in Fig. 6, which shows the density profiles from Fig. 1 ($p = 1$ torr) normalized to the central electron density. Substituting Eq. (12b) into Eq. (2) or (3), we find that, with a significant degree of electronegativity ($n_n > n_e$), the densities in the inner region satisfy the relationship

$$n_p(r) \sim n_n(r) \sim n_e(r) \sim J_0(r/l_0). \quad (13a)$$

For plane geometry, the Bessel function should be replaced with $\cos(x/l_0)$. In Eq. (13a), the characteristic length [10, 26]

$$l_0^2 = \frac{D_{an}}{\nu_d} + \frac{\nu_a D_{ap}}{\nu_i \nu_d} \approx \frac{\Lambda^2 \bar{n}_n}{\nu_i \tau_{ap} \bar{n}_e} > \Lambda^2 \quad (13b)$$

also determines the ambipolar electric field ($E(r) = -T_e \nabla n_e/n_e$) in the central region ($r < r_0$):

$$E^{(0)}(r) \approx -T_e J_1(r/l_0)/l_0 \approx -T_e r/l_0^2. \quad (14a)$$

Since the conditions $l_0 > \Lambda > l_e$ are usually satisfied, density profiles (13a) in the inner region are flatter than in the outer region, and, when they are extended up to the wall, they do not turn to zero (see Eq. (5)). Consequently, field (14a) is weaker than the electric field in the shell ($r_0 < r < R$), for which we have from Eq. (10) the following estimate:

$$E^{(1)}(r) \approx -\frac{\pi T_e}{2l_e}\cot\frac{\pi(R-r)}{2l_e} \approx -\frac{\pi^2 T_e}{2l_e^2}(r - r_0). \quad (14b)$$

To illustrate the limiting cases, we use Eqs. (2) and (3) to rewrite relationship (11a) in the form [10, 26]

$$\Gamma_n = \nu_a \int_{r_0}^{R} n_e(r)r dr = -\Gamma_p D = D\nu_i \int_{0}^{r_0} n_e(r)r dr, \quad (11b)$$

where $D = (D_n/D_p) \sim 1$. This relationship means that the number of ions that undergo attachment in the outer region is equal to the number of ions produced in the central region due to ionization. In the thin shell, a comparatively small flux of negative ions Γ_n is produced due to attachment; hence, a fairly weak electric field (14a) is sufficient to transport these ions into the inner region, in which they disappear due to detachment. Since, at $\tau_{an}\nu_a > 1$, the electrons in the inner region disappear mainly due to attachment, it is necessary to enable just a minor flux of positive ions toward the
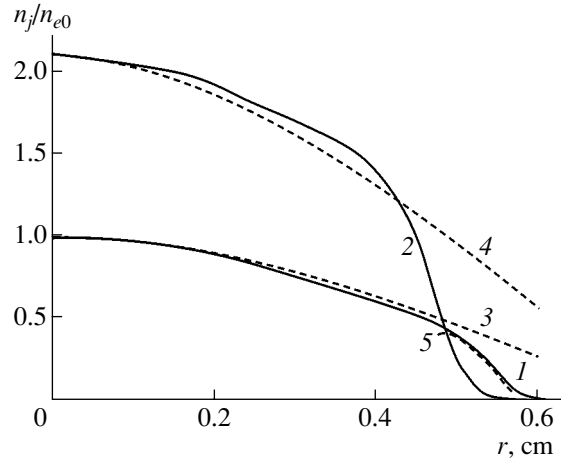
Fig. 6. Normalized density profiles for $p = 1$ torr and $I = 50$ mA: (1) $n_e(r)/n_{e0}$, (2) $n_n(r)/n_{e0}$. Curves *3* and *4* show the results calculated by formula (13), and curve *5* shows profile (10) in the outer region.

outer region. In other words, relationship (11b) means that, if the local plasmochemical balance of ions dominates over their spatial transport, the latter should only compensate for a relatively small difference between the attachment and detachment of negative ions.

Using expression (10) for $n_e(r)$ and Eqs. (14), we can obtain from Eq. (11b) the ionization frequency ν_i, which represents the eigenvalue of the boundary value problem described by Eqs. (2) and (3) [26]. The simple estimate $\Gamma_n \approx \nu_a n_e l_e \approx \Gamma_p \approx \nu_i n_e \Lambda$ gives $\nu_i \approx n_a l_e/\Lambda \approx \sqrt{\nu_a/\tau_{an}}$ [10, 26]. In the case at hand, we have $\nu_i \tau_{an} \approx \sqrt{\nu_a \tau_{an}} > 1$; hence, we obtain $\tau_{an}\nu_i > 1$. This means that the ionization frequency exceeds the value given by the Schottky formula for a simple plasma ($\tau_{an}\nu_i = 1$) [25, 26].

The density of the negative ions that are produced in the shell due to attachment can be deduced from their flux Γ_n (11b):

$$n_n \approx \Gamma_n/b_n E^{(1)}$$
$$\approx \frac{8l_e^2 n_e(r_0)}{\pi^2 D_{an}}\sin^2\frac{\pi(R-r)}{2l_e}\tan\frac{\pi(R-r)}{2l_e}. \quad (15)$$

This density is much lower than the densities of electrons and positive ions,

$$n_n \approx \frac{\pi n_e(r_0)\nu_a}{4D_{an}l_e}(r - R)^3 \quad (r_0 < r \le R). \quad (15a)$$

At the point $r = r_0 \approx R - l_e$, the field $E^{(1)}$ is close to zero, whereas the flux Γ_n (11b), caused by attachment in the outer region, is finite. Therefore, when approaching the point $r = r_0$, negative ion density (15) sharply

increases,

$$n_n \approx \frac{8l_e^3 n_e(r_0)\nu_a}{\pi^3 D_{an}(r-r_0)} \quad (r \ge r_0 = R - l_e), \qquad (15b)$$

to its value in the inner region, which is determined by Eq. (12). The transition zone separating regions *0* and *1* with different ion compositions is narrow $(\sim l_n < l_e)$. For this reason, it was treated in [10, 26] as a diffusive jump in which ion densities undergo a break, whereas the ion fluxes and the electron density are continuous. The validity of relationship (12) in the region $r < r_0$ in an oxygen discharge is illustrated in Fig. 6, which shows the normalized density profiles obtained from the profiles in Fig. 1 at a pressure of $p = 1$ torr. The dashed curves in Fig. 6 show the profiles calculated by formula (13) for the inner region and by formula (10) for the shell with the thickness $R - r_0 \approx l_e$. When deducing formula (10) for the outer region, the shell thickness δ_{sh} was taken into account; i.e., it was assumed that the electron density vanished at $r = R - \delta_{sh}$, rather than at the tube wall. It can be seen that the results of calculations by these formulas agree well with the results of full-scale simulations.[1]

At lower pressures, the role of spatial transport increases and, thus, the characteristic length l_e (8) and l_n (9) also increase. The increase in length l_0 (13b) leads to the flattening of density profiles (13a) in the inner region. Because of the increase in the length l_n, the region with a sharp change of the ion density spreads out due to ion diffusion; hence, the transient region can no longer be treated as a jump. As a result, the ion density profiles become bell-shaped.

At $l_n \ge \Lambda$, the negative ions are able to pass throughout the entire discharge volume due to their diffusion. However, they turn out to be trapped in the inner region by the electric field; as a result, a Boltzmann distribution (similar to that for electrons) is established:

$$-T_e \nabla n_e/n_e = -T \nabla n_n/n_n = E. \qquad (16)$$

It follows from Eqs. (8) and (9) that, generally, the self-diffusion of negative ions prevails $(l_n > \Lambda)$ only when attachment is insignificant as compared to the ambipolar diffusion of negative ions (ion diffusion with the electron temperature), i.e., when $\tau_{an}\nu_a \ll 1$ (see [24–26] for details).

Condition (16) leads to the relationship

$$n_e(r)/n_e(0) = [n_n(r)/n_n(0)]^{1/k}, \qquad (17)$$

which strongly depends on the temperature ratio $k = T_e/T_i$ and coincides with distribution (12b) only in the particular case $T_e = T_i$. The establishment of a Boltzmann distribution for electrons and negative ions at low pressures is illustrated in Fig. 7, in which the simulation results shown in Fig. 3 for a pressure of $p = 0.15$ torr are replotted in accordance with Eq. (17).

Since $k \gg 1$ in the discharge, it follows from Eq. (17) that the electron density profile is nearly flat, $n_e(r) \approx n_{e0} \approx$ const, which is indeed observed at reduced pressures (see Figs. 2, 3). Here, transport processes play a major role in Eq. (3) for the negative ion balance (see Fig. 8), in contrast to the above case with $\tau_{an}\nu_a > 1$ (cf. Fig. 5). The field-induced and diffusion fluxes of negative ions are almost the same in magnitude, but opposite in sign; hence, a small difference between them is sufficient to balance the production and loss of ions at any point (Fig. 8). The plasmochemical processes govern only the global balance of ions in the central region. In Eq. (2) for the positive ion density $n_p(x)$, the terms on the left-hand side are also approximately equal to each other. However, they are summed up and, thus, at a significant degree of electronegativity ($n_n(0) > n_e(0)$), balance equation (2) for positive ions can be written in the form $-2D_p\Delta n_n = \nu_i n_{e0}$. This gives a parabolic distribution of the ion densities and a flat profile of the electron density $n_e(r)$ at $r < r_0$ [13–15]:

$$n_n(r) = n_{n0}(1 - r^2/r_0^2),$$
$$n_{n0}/n_{e0} = \nu_i r_0^2/4D_p, \qquad (18)$$
$$n_e(r) \approx n_{e0} \approx \text{const}.$$

We note that ion diffusion in the inner region proceeds with the coefficient $2Dp$ of the own ion–ion ambipolar self-diffusion, rather than with the usual coefficient of ambipolar diffusion $D_p(1 + k)$. It can be seen from Figs. 2 and 3 that, at low pressures, simple parabolic law (18) for the ion density profiles agrees well with the results of full-scale simulations.[2]

In the outer region (shell), in which the negative ions are almost absent, the plasma density profile varies in accordance with Eq. (10). In [13–15, 23, 24], the position of the boundary point $r = r_0$ was found from the negative ion balance using model profiles (18). Unfortunately, this procedure is rather laborious and provides a low accuracy. It seems that the position of the boundary can be found in a simpler and more reliable way from the continuity of the positive ion flux at $r = r_0$:

$$2D_p\frac{n_{n0}}{r_0} = \frac{D_p(1+k)n_{e0}}{l_e\tan((R-r_0)/l_e)} \approx \frac{D_p(1+k)n_{e0}}{R-r_0}. \quad (19)$$

---

[1] Note that, for the recombination regime $(\nu_d < n_p K_r)$, it follows from Eqs. (11a) and (12a) that $\nabla n_e/n_e = \nabla n_n/n_n + \nabla n_p/n_p \approx 2\nabla n_n/n_n$, which results, in contrast to Eq. (12b), in an ion distribution that is flatter than the electron distribution (see [24–26] for details). In such a situation (which occurs, e.g., for halogens), the attachment and ionization frequencies are approximately the same, $\nu_i \approx \nu_a$, as was noted in [16–18].

[2] We note also that, in order for profiles (18) to be established, it is enough to satisfy the condition $\tau_{an}\nu_a < 1$. The mechanism for the volume loss of negative ions, which is determined by the right-hand side of Eq. (3), can be either recombination (at $\nu_d < K_r n_p$) or detachment (at $\nu_d > K_r n_p$).
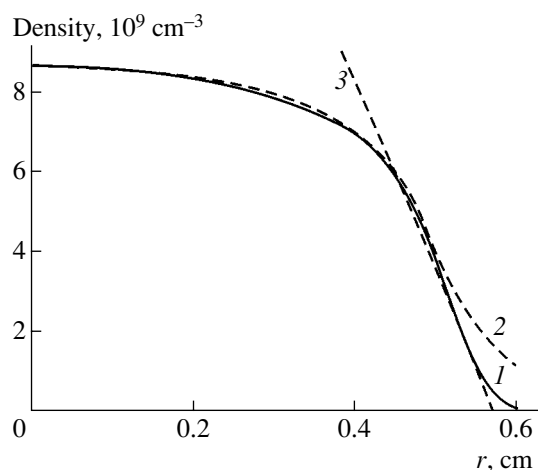
Density, 10^9 cm^{-3}

Fig. 7. Boltzmann distributions of electrons and negative ions for $p = 0.15$ torr and $I = 50$ mA: (*1*) $n_e(r)$, (*2*) $n_e(0)/[n_n(r)/n_n(0)]^{1/k}$ (see formula (17)), and (*3*) electron density profile (10) in the outer region.

Flux, 10^{15} cm^{-2}

Fig. 8. Contributions of spatial transport and volume processes to the negative ion balance for $p = 0.15$ torr and $I = 50$ mA: (*1*) the diffusion component of the negative ion flux, (*2*) its drift component, and (*3*) the resulting flux balancing the production and loss of negative ions in volume plasmochemical processes.

Model electron density profiles (10) with r_0 defined by Eq. (19) (see Fig. 7) agree well with the results of full-scale simulations shown in Fig. 3.

Based on the analysis performed, we recommend the following procedure to obtain approximate density profiles in the plasma of electronegative gases in the detachment regime ($\nu_d > K_r n_p$):

(i) First, the parameter $\tau_{an}\nu_a$ is estimated.

(ii) Then it is necessary to indent from the wall by the thickness δ_{sh} of the space charge sheath, which can be estimated, e.g., according to [26].

(iii) In the outer electron–ion plasma region ($r_0 < r < R$), where $n_p \approx n_e \gg n_n \approx 0$, the electron density varies according to Eq. (10) and the negative ion density varies according to Eq. (15). If $\tau_{an}\nu_a > 1$, then the thickness of this region is equal to l_e (see Eq. (8)) and the denominator in Eq. (10) is equal to unity ($r_0 = R - l_e$). In the opposite case ($\tau_{an}\nu_a < 1$), we have $l_e \approx \Lambda$ and the thickness of this region is estimated by formula (19).

(iv) Finally, the density profiles in the central region ($r < r_0$) are determined.

At $\tau_{an}\nu_a > 1$, the density profiles are similar and are described by Eq. (13), whereas the density values are related by expression (12). Electron density profile (13) is joined to expression (10) at $r = r_0 = R - l_e$. The ion densities undergo a jump at this point: the negative ion density drops to nearly zero (see Eq. (15)), whereas the positive ion density decreases to the value equal to the electron density given by Eq. (10). At $\tau_{an}\nu_a > 1$, the thickness of the transition zone ($\sim l_n < l_e$) is small and it can be regarded as a jump in the ion density.

At $\tau_{an}\nu_a < 1$, the electron density profile is flat ($n_e(x) \approx n_{e0}$) and the ion density profile is parabolic. These densities are related by formulas (18). The elec-
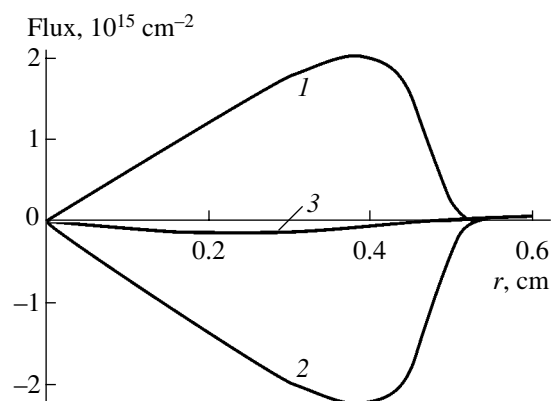
tron density profile is joined at the point $r = r_0$, whose position can be estimated from Eq. (19).

Thus, using commercial CFDRC software [28], comprehensive simulations of the positive column plasma of a dc discharge in oxygen are performed and the main scaling laws characterizing the spatial distributions of the plasma parameters are determined. The simulation results show that a distinctive feature of the electronegative-gas plasma is that it stratifies into regions with very different ion compositions, so that there are practically no negative ions in the outer electron–ion plasma region (shell). At low pressures ($\tau_{an}\nu_a < 1$), not only electrons but also negative ions obey a Boltzmann distribution. In the inner region, the electron density profile is flat, whereas the ion density profile is parabolic. In the ion balance, the transport processes prevail; hence, taking into account ion heating dramatically affects the spatial distribution of charged particles. At elevated pressures ($\tau_{an}\nu_a > 1$), the volume processes dominate in the balance of negative ions and the profiles of the charged particle densities in the inner region become similar to each other.

ACKNOWLEDGMENTS

REFERENCES

1. M. Lieberman and A. Lichtenberg, *Principles of Plasma Discharges and Materials Processing* (Wiley, New York, 1994).

2. H. J. Oskam, Philips Res. Rep. **13**, 335 (1958).

3. M. V. Konjukov, Zh. Éksp. Teor. Fiz. **34**, 908 (1958) [Sov. Phys. JETP **7**, 629 (1958)]; Zh. Éksp. Teor. Fiz. **34**, 1634 (1958) [Sov. Phys. JETP **7**, 1122 (1958)].

4. J. B. Tompson, Proc. Phys. Soc. (London) **73**, 818 (1959).

5. H. Sabadil, Beitr. Plasmaphys. **13**, 235 (1973).

6. H. S. W. Massey, *Negative Ions* (Cambridge University Press, Cambridge, 1976; Mir, Moscow, 1979).

7. P. D. Edgeley and A. von Engel, Proc. R. Soc. London, Ser. A **370**, 375 (1980).

8. G. L. Rogoff, J. Phys. D **18**, 1533 (1985).

9. A. V. Phelps, J. Res. Natl. Inst. Stand. Technol. **95**, 407 (1990).

10. L. D. Tsendin, Zh. Tekh. Fiz. **55**, 2318 (1985) [Sov. Phys. Tech. Phys. **30**, 1377 (1985)]; Zh. Tekh. Fiz. **59**, 21 (1989) [Sov. Phys. Tech. Phys. **34**, 11 (1989)].

11. C. M. Ferreira, G. Gousset, and M. Touzeau, J. Phys. D **21**, 1403 (1988).

12. P. R. Daniels, R. N. Franklin, and J. Snell, J. Phys. D **22**, 780 (1989); J. Phys. D **23**, 823 (1990); J. Phys. D **26**, 1636 (1993).

13. A. J. Lichtenberg, V. Vahedi, M. A. Lieberman, *et al.*, J. Appl. Phys. **75**, 2339 (1994).

14. A. J. Lichtenberg, I. G. Kouznetsov, T. D. Lee, *et al.*, Plasma Sources Sci. Technol. **6**, 437 (1997).

15. I. G. Kouznetsov, A. J. Lichtenberg, and M. A. Lieberman, J. Appl. Phys. **86**, 4142 (1999).

16. R. N. Franklin, Plasma Sources Sci. Technol. **11**, 31 (2002).

17. R. N. Franklin and J. Snell, J. Phys. D **32**, 2190 (1999).

18. R. N. Franklin, J. Plasma Phys. **64**, 131 (2000).

19. S. V. Berezhnoj, C. B. Shin, U. Buddemeier, and I. Kaganovich, Appl. Phys. Lett. **77**, 800 (2000).

20. V. A. Feoktistov, V. V. Ivanov, A. M. Popov, *et al.*, J. Phys. D **30**, 423 (1997).

21. V. V. Ivanov, K. S. Klopovsky, D. V. Lopaev, *et al.*, IEEE Trans. Plasma Sci. **27**, 1279 (1999).

22. V. V. Ivanov, K. S. Klopovsky, D. V. Lopaev, *et al.*, Fiz. Plazmy **26**, 1038 (2000) [Plasma Phys. Rep. **26**, 972 (2000)].

23. A. Kono, Appl. Surf. Sci. **192**, 115 (2002).

24. E. A. Bogdanov and A. A. Kudryavtsev, Pis'ma Zh. Tekh. Fiz. **27** (21), 36 (2001) [Tech. Phys. Lett. **27**, 905 (2001)].

25. E. A. Bogdanov, V. I. Kolobov, A. A. Kudryavtsev, and L. D. Tsendin, Zh. Tekh. Fiz. **72** (8), 13 (2002) [Tech. Phys. **47**, 946 (2002)].

26. A. V. Rozhansky and L. D. Tsendin, *Transport Phenomena in Partially Ionized Plasma* (Énergoatomizdat, Moscow, 1988; Taylor & Francis, London, 2001).

27. E. A. Bogdanov, A. A. Kudryavtsev, L. D. Tsendin, *et al.*, Zh. Tekh. Fiz. **73** (8), 45 (2003) [Tech. Phys. **48**, 983 (2003)].

28. http://www.cfdrc.com/~cfdplasma

29. E. W. McDaniel and E. Mason, *The Mobility and Diffusion of Ions in Gases* (Wiley, New York, 1973; Mir, Moscow, 1976).

Translated by N. Ustinovskiĭ

===================== **SOLIDS** =====================

Magnetoresistive Effect in $RCu_3Mn_4O_{12}$ Perovskite-Like Oxides

L. S. Lobanovsky, I. O. Troyanchuk, S. V. Trukhanov, S. N. Pastushonok, and V. I. Pavlov

Institute of Solid-State and Semiconductor Physics, Belarussian Academy of Sciences,
ul. Brovki 17, Minsk, 220072 Belarus
e-mail: Lobanov@ifttp.bas-net.by
Received November 4, 2002; in final form, February 26, 2003

Abstract—The electrical properties of and the magnetoresistive effect in $RCu_3Mn_4O_{12}$ (R = rare-earth ion or Th) are studied. In all compounds of this series, the magnetoresistive effect amounts to 20% at liquid nitrogen temperature in the presence of a field of 0.9 T. An increase in the magnetoresistance with decreasing temperature and a high sensitivity to weak magnetic fields at low temperatures point to the intergranular nature of the effect. The magnetoresistance shows a peak in the vicinity of the Curie temperature T_C. Based on the dependences of the magnetoresistance on an external magnetic field, it is assumed that the magnetoresistance peak near T_C is related to the charge carrier scattering by magnetic inhomogeneities as in substituted orthomanganites. We believe that the magnetoresistance value near the magnetic ordering temperature depends on the synthesis conditions and the effect of the intergranular spacer on the transport properties of these compounds. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The discovery of colossal magnetoresistance in perovskite-like manganites [1, 2] has stimulated the search for new compounds exhibiting this effect because of its great practical importance [3]. In magnetically ordered systems, the maximum of magnetoresistance at the magnetic order–disorder transition is related to the scattering of charge carriers by magnetic inhomogeneities. The value of the magnetoresistance depends on the electrical conductivity jump near T_C and an increase in the Curie temperature in the presence of a magnetic field (i.e., on the sensitivity of the magnetic order parameter to an external magnetic field). The Magnetoresistance peak near T_C is observed in $La_{1-x}(Ca, Sr)_xMnO_3$ (LCSMO) substituted lanthanum orthomanganites with a perovskite structure [1, 2, 4, 5]. Near the magnetic ordering temperature, substituted lanthanum orthomanganites also exhibit the metal–insulator transition [5].

Materials exhibiting colossal magnetoresistance are used in data processing and storage devices. However, the above features are observed in strong magnetic fields, which limits the application of the magnetoresistive effect.

In addition to the magnetoresistance observed at the metamagnetic transition and the scattering of charge carriers by magnetic inhomogeneities near the Curie temperature, there exists another type of magnetoresistive effect. Giant magnetoresistance (GMR) was discovered in grained magnetic materials with a high spin polarization of charge carriers [6, 7]. The GMR increases monotonically with decreasing temperature and is highly sensitive to weak magnetic fields. This effect shows up most vividly at temperatures significantly lower than the magnetic ordering temperature. The GMR phenomenon has not yet clearly understood, and the value of the magnetoresistance for a specific compound cannot be predicted theoretically. However, in the simple model of electron jumps over an insulating barrier (intergranular spacer in our case), the magnetoresistance is represented as $MR = \Delta\rho/\rho = -(JP/4kT)[M^2(H, T) - M^2(0, T)]$, where J, P, and M are the exchange interaction constant, polarization of tunneling electrons, and magnetization of the material, respectively. The GMR behavior in grained nickel films [8, 9], polycrystalline Fe_3O_4 films [10], and $La_{2/3}Sr_{1/3}MnO_3$ perovskite [11] poorly agrees with the theoretical predictions. Kobayashi *et al.* [12] discovered an intergranular magnetoresistive effect in ferrimagnetic double perovskites $Sr_2Fe(Mo$ or $Re)O_6$ with a composition-dependent Curie temperature between 410 and 450 K. As was demonstrated in [12], the *MR* behavior in these compounds shows the best agreement with predictions based on the model of spin-polarized charge carrier scattering by interfaces between neighboring granules with opposing magnetizations. However, the discrepancy between the temperature dependences of the magnetization and magnetoresistance still remains to be explained. Therefore, the search for model magnetic systems with a high spin polarization of carriers is an important problem.

Earlier, Troyanchuk *et al.* [13, 14] studied the electrical and magnetic properties of $CaCu_{1-x}Mn_xMn_4O_{12}$ and $RECu_3Mn_4O_{12}$ (RE = Tb, Tm) systems, which exhibit the GMR effect, and $TmCu_3Mn_4O_{12}$ and $Ca(Cu_{1.5}Mn_{1.5})Mn_4O_{12}$ systems, where the metal–insulator transition is observed. However, the temperature of the transition differs greatly from the magnetic ordering temperature. As was mentioned, there is a correlation between T_C and the metal–insulator transition in LCSMO systems. The peak of the magnetoresistance near the Curie temperature in $CaCu_{1-x}Mn_xMn_4O_{12}$ ($0 \le x \le 3$) and $RECu_3Mn_4O_{12}$ (RE = Tb, Tm) solid solutions was not observed.

A comprehensive study of the electrical and magnetic properties of $RCu_3Mn_4O_{12}$ (R = rare-earth ion or Th) compounds has not been carried out. Theoretically, it is of interest to investigate the temperature dependence of the magnetoresistance and metal–insulator transition, which probably correlates with the magnetic ordering temperature as in manganese-containing perovskite-like oxides. From the practical point of view, it is important to measure the magnetoresistance of compounds with a high spontaneous magnetization, such as $RCu_3Mn_4O_{12}$, since the theory predicts the quadratic magnetization dependence of the magnetoresistance for temperatures much below T_C.

EXPERIMENTAL

$RCu_3Mn_4O_{12}$ samples were made of related oxides. Extra-pure-grade $R_2^{3+}O_3^{2-}$, $R^{4+}O_2^{2-}$ $Cu^{2+}O^{2-}$, $Mn_2^{3+}O_3^{2-}$, and $Mn^{4+}O_2^{2-}$ oxides taken in stoichiometric proportions were ground in an alundum mortar. The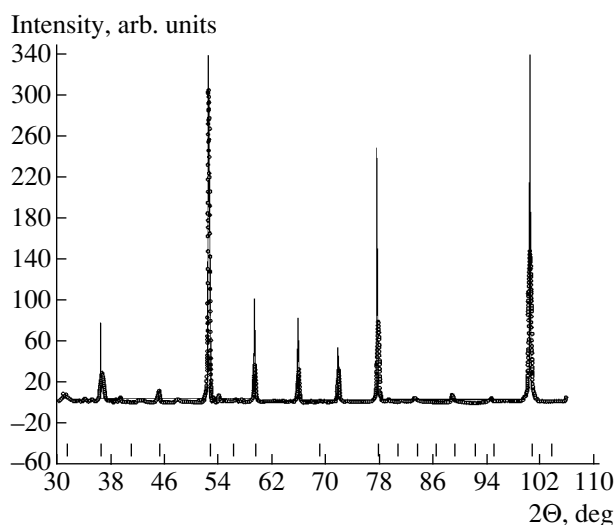 weights of the components were taken according to the rare-earth ion valence. Prior to synthesis, the charge was pressed under a pressure of 0.1 GPa to form cylindrical pellets of diameter 10 mm and height 8–10 mm. Then, the pellets were covered by nickel foil and placed in the graphite heater of a high-pressure container. The foil isolates the sample during the synthesis from the graphite heater, which is an effective reducer. The samples were prepared by solid-phase synthesis proceeding under a high pressure and high temperature. A lithographic stone served as a pressure-transmitting medium. The synthesis took 10 min at a pressure of 5 GPa and a temperature of about 1670 K. The exact value of the temperature was varied with rare-earth ions.

With the X-ray diffraction technique (DRON-3 diffractometer), we measured the homogeneity of the compounds synthesized and refined the parameters of the unit cell.

The magnetization was measured with a vibrating-sample magnetometer at various temperatures. The dynamic magnetic susceptibility was measured using a mutual-inductance bridge; the electrical properties and magnetoresistance, with a conventional four-probe technique. The magnetoresistive effect $MR = \Delta\rho/\rho = (\rho_{H=0} - \rho_H)/\rho_{H=0} \times 100\%$ was estimated in an external magnetic field of 0.9 T.

RESULTS AND DISCUSSION

Crystal lattice. All the compounds $RCu_3Mn_4O_{12}$ have cubic symmetry of the unit cell with a doubled parameter a of the ABO_3 perovskite structure (Fig. 1). The doubling of the unit cell parameter results from the fact that rare-earth and copper ions arrange in the ratio 1 : 3 and also because oxygen octahedrons are tilted to $\langle 100 \rangle$ axes. The Mn–O–Mn bond angle along $\langle 100 \rangle$ directions becomes equal to 140° instead of 180°, which is typical of the undistorted perovskite structure. The X-ray diffraction patterns show weak reflections from impurity phases. The relative impurity concentration was no greater than 1.5%. The unit cell parameters were refined in view of the fact that $CaCu_3Mn_4O_{12}$ belongs to the *Im*3 space group [15]. The table lists the unit cell parameters of the solid solutions under study. The unit cell volume is seen to decrease insignificantly with increasing atomic number of the rare earth, which is in agreement with the decrease in the ion radius of the rare earth at the constant valent state of the remaining ions. The unit cell volume of the compounds containing tetravalent cerium and thorium ions is greater than that of the compounds with the trivalent rare-earth ion in the A position. Such behavior is related to the increase in the concentration of trivalent manganese ions from 25% (trivalent ion in the A position) to 50% (tetravalent ion). It is known [16] that the ion radius of Mn^{3+} is greater than that of Mn^{4+}.

Magnetic properties. The measurements of the magnetization in the external magnetic field at low tem-

Fig. 1. X-ray diffraction pattern (CrK_α radiation) of the compound $CeCu_3Mn_4O_{12}$. The solid line shows the theoretical spectrum calculated using the refined unit cell parameter. Note the positions of the Bragg angles.

peratures show that all the compounds of the given series have a spontaneous magnetic moment. A maximum moment of 9.7 Bohr magnetons per formula unit (μ_B/f.u.) is reached at 10 K in the thorium-containing compound (see table). Neutron diffraction experiments [17] show that the magnetic moments of manganese and copper ions in $AC_3B_4O_{12}$ compounds are ordered oppositely. The magnetic moment $\mu(Cu^{2+})$ of a divalent copper ion is 1 μ_B, whereas the magnetic moments of tri- and tetravalent manganese ions are $\mu(Mn^{3+}) = 3.5\ \mu_B$ and $\mu(Mn^{4+}) = 2.6\ \mu_B$ [17]. Therefore, we expect the $R^{3+}Cu^{2+}{}_3Mn^{3+}{}_3Mn^{4+}O^{2-}_{12}$ electron configuration with the trivalent rare-earth ion in the A position. When the magnetic moments $\mu(3Cu^{2+}) = 3\mu_B$ and $\mu(Mn^{3+}{}_3Mn^{4+}) = 11.1\mu_B$ are antiparallel, the magnetization is 8.1 μ_B per formula unit. For compounds with the paramagnetic ion in the A position, the peak in the temperature dependence of the magnetization is related to the antiparallel ordering of the rare-earth ion magnetic moments relative to the total magnetic moment of the A and C sublattices. The samarium compound has the maximum temperature (75 K) of magnetic moment ordering.

The difference between the theoretical and experimental values of the magnetizations is presumably associated with trivalent manganese ions partially substituting for copper ions, since the C positions of these compounds can be occupied only by Jahn–Teller ions. The magnetic moment of the copper ion is opposite to that of the manganese sublattice, whereas the magnetic moment of the Mn^{3+} ion substituting for copper is codirected with the total magnetic moment of the manganese sublattice. Based on the magnetic data, it is difficult to judge the cation distribution in these compounds. However, the Curie temperature of the solid solutions obtained at a high pressure (Fig. 2) differs only slightly from the magnetic ordering temperature of the compounds synthesized by the hydrothermal method ($T_C = 390$ K) [17]. This means that the real chemical composition is close to the chemical formula.

Electrical properties and magnetoresistance. The electrical conductivity of substituted orthomanganites with a perovskite structure depends on the presence of manganese ions with different valences in one sublattice [18]. The conduction band of these compounds forms when the partially filled d subshell of manganese ions and the p subshell of oxygen ions overlap. The width of the conduction band depends on the Mn–O–Mn bond length and angle. As was mentioned, the Mn–O–Mn bond angle in $RCu_3Mn_4O_{12}$ compounds (about 140°) depends on the rare-earth ion in the A position insignificantly. It is likely that, such behavior is due to the presence of Jahn–Teller Cu^{2+} and Mn^{3+} ions in the C position.

Figures 3a and 4a show the temperature dependence of the resistivity of $RCu_3(Mn,Cr)_4O_{12}$ (R = Ce, Sm, Ho,

Unit cell parameters and spontaneous magnetic moments of $RCu_3Mn_4O_{12}$ compounds

R	a, Å	M, μ_3/f.u.
Ce^{4+}	7.349	6.4
Sm^{3+}	7.284	7.8
Gd^{3+}	7.265	2.9
Tb^{3+}	7.258	5.6
Ho^{3+}	7.253	5.0
Tm^{3+}	7.247	–
Yb^{3+}	7.246	9.5
Th^{4+}	7.383	9.7

Yb, Th) compounds. Despite the equal ratio of the manganese ions with different valences, $RCu_3Mn_4O_{12}$ (R = Ce^{4+}, Th^{4+}) compounds exhibit different temperature dependences of the resistivity. The conductivity of the $ThCu_3Mn_4O_{12}$ solid solution decreases with temperature, which is characteristic of semiconductors. The cerium composition has metallic conductivity at temperatures below the magnetic ordering point. Similarly to substituted lanthanum orthomanganites, the resistivity of this compound exhibits semiconductor behavior at temperatures above T_C. Substituting chromium ions for manganese ions in the B position leads to a sharp increase in the resistivity and decreases the Curie temperature (Fig. 2). In addition, this solid solution acquires semiconductor properties throughout the temperature range. The temperature dependences of the resistivity for the samarium-, ytterbium-, holmium-, and thulium-based compounds peak at temperatures other than the magnetic ordering temperature (Fig. 4a). The resistivity of $TmCu_3Mn_4O_{12}$ compounds behaves in a similar way [14]. The resistivity curves of the samarium- (inset in Fig. 4) and ytterbium-based compounds exhibit a metal–insulator transition and bends
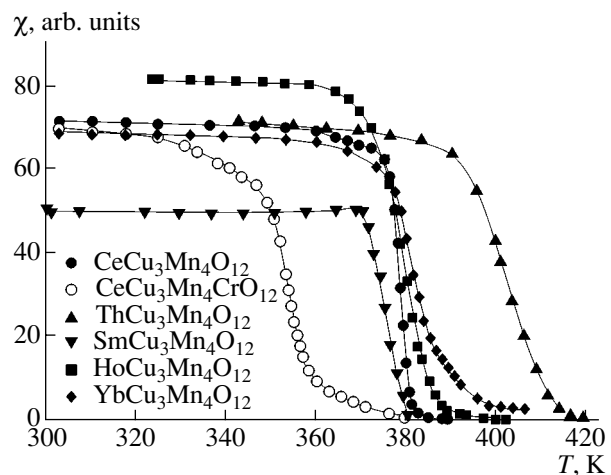

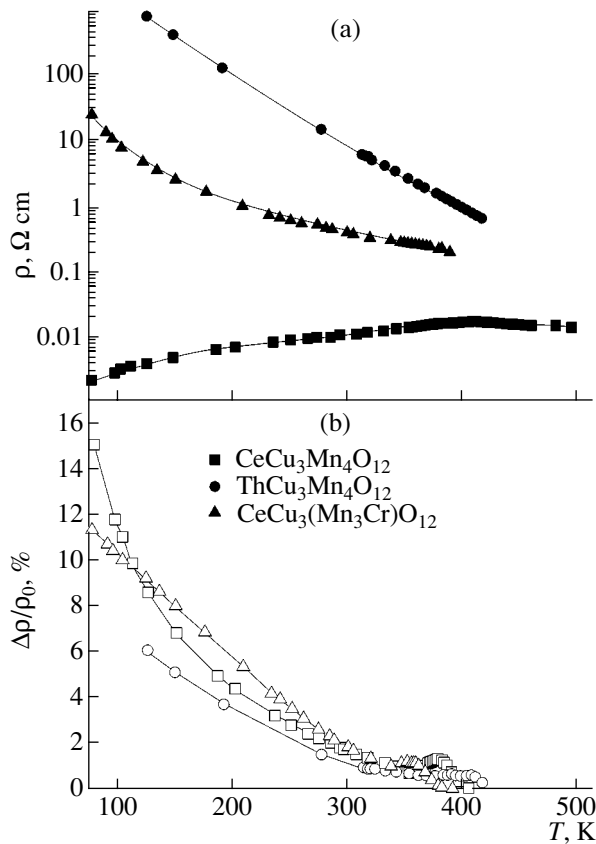**Fig. 2.** Dynamic magnetic susceptibility of $RCu_3Mn_4O_{12}$ compounds versus temperature.

Fig. 3. (a) Resistivity of and (b) magnetoresistive effect in the compounds $CeCu_3(Mn,Cr)_4O_{12}$ and $ThCu_3Mn_4O_{12}$ with the tetravalent ion in the *A* position versus temperature.

Fig. 4. (a) Resistivity of and (b) magnetoresistive effect in $RCu_3Mn_4O_{12}$ (R = Sm, Ho, Yb) compounds with the trivalent ion in the *A* position versus temperature. The arrows indicate the temperatures of the metal–insulator transition. The inset shows the behavior of the magnetoresistance near the magnetic ordering temperature.

related to the Curie temperature. The metal–insulator transition for the ytterbium compound is strongly diffused. For this compound, the semiconductor character of the temperature dependence of the resistivity is most pronounced at temperatures exceeding T_C.

In all the $RCu_3Mn_4O_{12}$ compounds, the magnetoresistive effect is observed at temperatures below the temperature of the magnetic order–disorder transition (Figs. 3b, 4b). The maximum value of the magnetoresistance (18%) is reached in the ytterbium compound at the lowest temperature of our experiments (78 K). The amount of the effect decreases monotonically with growing temperature. The theory predicts that the effect vanishes at the Curie temperature. However, the magnetoresistance curves for all the compounds run anomalously near T_C. In particular, the curves of the cerium, holmium, and samarium compounds show peaks, the highest of which (2.4%) is observed in $SmCu_3Mn_4O_{12}$ at a field strength of 0.9 T. Substituting chromium ions for manganese ions in $CeCu_3(CrMn_3)O_{12}$ reduces insignificantly the low-temperature magnetoresistance and does not affect the anomalous behavior of the magne-

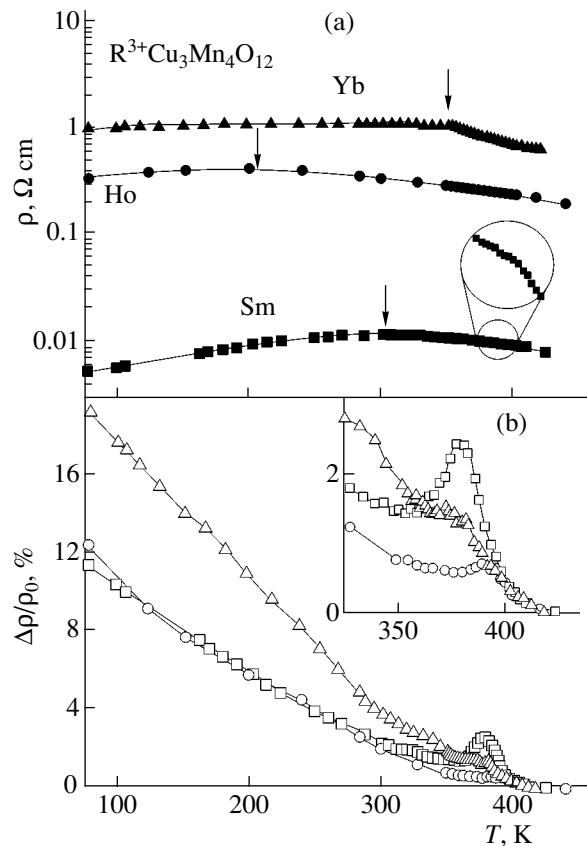toresistance near the Curie temperature. This points to the different nature of the magnetoresistance at various temperatures. To shed light on the origin of the effect, we measured the field dependences of the magnetoresistance at low temperatures and at the Curie temperature (Fig. 5). It is seen that, at low temperatures, the magnetoresistance saturates in weak magnetic fields and varies only slightly as the field grows. Near T_C, the magnetoresistive effect depends quasilinearly on the external field and does not saturate in fields up to 1 T. Both the temperature and field dependences of the magnetoresistance suggest the intergranular character of the phenomenon at low temperatures. It is assumed that the magnetoresistance peak near the Curie point is related to the scattering of charge carriers by magnetic inhomogeneities. Similar behavior is observed for substituted lanthanum orthomanganites and Ba_2FeMoO_6 double perovskite [19]. However, in LCSMO systems, the magnetoresistance peak is observed at a Mn^{4+} ion concentration of 20–45%. In this concentration range, substituted lanthanum manganites are ferromagnets at temperatures below the magnetic ordering point [1, 2, 5], whereas in the ferrimagnetic compounds under

study, the concentrations of Mn^{4+} ions are 50 and 75% for the tetra- and trivalent ion in the *A* position, respectively.

Note that the magnetoresistance peak (2.4%) observed for the samarium compound at 375 K is the highest of the peaks observed for all the $RCu_3Mn_4O_{12}$ compounds near the magnetic ordering temperature. The magnetoresistive of $Ba_2(FeMo)O_6$ double perovskite ($T_C = 375$ K) at a field strength of 1 T has the same value at the Curie temperature.

In our opinion, the metal–insulator transition in the compounds under study is related to the presence of an intergranular layer (spacer) and the high content of defects in samples prepared at a high pressure. Apparently, the properties of the spacer govern the electrical properties and the metal–insulator transition temperature in $RCu_3Mn_4O_{12}$ compounds. As was demonstrated [20] with $La_{0.82}Sr_{1.08}MnO_z$ polycrystalline solid solution ($z \approx 3$), the metal–insulator transition, which is typical of substituted lanthanum manganites at the Curie temperature, shifts towards lower temperatures, because the spacer affects the magnetic and electrical properties of this compound. Weak exchange interaction between magnetic ions in the spacer reduces the temperature of transition to the paramagnetic state and that of change of the conduction type compared to these temperatures inside grains with regular crystal and magnetic order. The X-ray and electrical data in combination support the assumption that the spacer has an effect on the electrical conductivity of the $RCu_3Mn_4O_{12}$ compounds. The cerium and ytterbium compounds have the narrowest spectral bands. The bands become successively broader in the samarium and holmium compounds. According to the Scherrer relationship [21], spectral lines broaden with a decrease in the grain size in a polycrystal. The temperature of the metal–insulator transition decreases in the same sequence of compounds. Similarly to LCSMO compounds, the effect of grains on the electrical properties of $SmCu_3Mn_4O_{12}$ and $YbCu_3Mn_4O_{12}$ is noticeable at T_C. For these compounds, we observe a sharp change in the activation energy in the resistivity curve.

We believe that the difference between the electrical properties of the cerium and thorium compounds is related to the critical length of the Mn–O–Mn bond. An increase in the unit cell parameter of $ThCu_3Mn_4O_{12}$ results in a deeper localization of carriers and, consequently, in the change of the conduction type. Based on this assumption, one can explain the electrical properties of $CeCu_3(CrMn_3)O_{12}$. The incorporation of Cr^{3+} ions into the manganese sublattice also causes a deep localization of charge carriers, since the ion radius of Cr^{3+} is smaller than that of Mn^{3+}.

The optimization of the $RCu_3Mn_4O_{12}$ synthesis conditions is expected to substantially enhance the magnetoresistive effect near the Curie temperature. $RCu_3Mn_4O_{12}$ compounds are magnetoresistive materi-
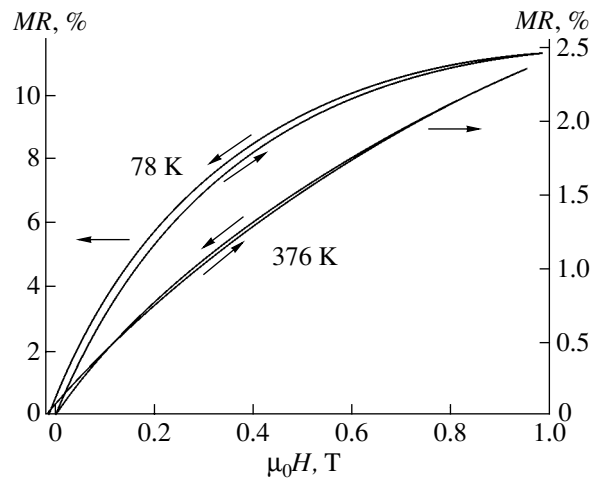
Fig. 5. Magnetoresistive effect in the compound $SmCu_3Mn_4O_{12}$ versus the magnetic field at two temperatures.

als that are promising for the room-temperature operation of magnetic heads or other devices based on magnetic-to-electric signal conversion.

In perovskite-like $RCu_3Mn_4O_{12}$ (R = rare-earth ion or Th) oxides, the magnetoresistive effect consists of two components: intergranular, which is the most pronounced at low temperatures and is related to the scattering of spin-polarized charge carriers at the interfaces between neighboring granules, and intragranular, which appears near T_C and is due to the scattering of carriers by magnetic inhomogeneities inside the grains. In this study, the Curie temperatures of all the compounds are above room temperature. The properties of the spacer considerably affects the conductivity of the solid solutions. It seems that the magnetoresistance may be enhanced, e.g., by optimizing the properties of the intergranular and intragranular layers.

ACKNOWLEDGMENTS

REFERENCES

1. J. Volger, Physica (Utrecht) **20**, 49 (1954).

2. S. Jin, T. H. Tiefel, M. McCormack, *et al.*, Science **264**, 413 (1994).

3. L. l. Balcells, A. E. Carrillo, B. Martínez, *et al.*, J. Magn. Magn. Mater. **221**, 224 (2000).

4. A. Anane, C. Dupas, K. Le Dang, *et al.*, J. Phys.: Condens. Matter **7**, 7015 (1995).

5. J. Fontcuberta, B. Martínez, A. Seffar, *et al.*, Phys. Rev. Lett. **76**, 1122 (1996).

6. A. Milner, A. Gerber, B. Groisman, *et al.*, Phys. Rev. Lett. **76**, 475 (1996).

7. J.-H. Park, E. Vescovo, H.-J. Kim, *et al.*, Nature (London) **392**, 794 (1998).

8. J. I. Gittleman, Y. Goldstein, and S. Bozowski, Phys. Rev. B **5**, 3609 (1972).

9. J. S. Helman and B. Abeles, Phys. Rev. Lett. **37**, 1429 (1976).

10. J. M. D. Coey, A. E. Berkowitz, L. I. Balcels, *et al.*, Appl. Phys. Lett. **72**, 734 (1998).

11. H. Y. Hwang, S.-W. Cheong, N. P. Ong, *et al.*, Phys. Rev. Lett. **77**, 2041 (1996).

12. K.-I. Kobayashi, T. Kimura, H. Sawada, *et al.*, Nature (London) **395**, 677 (1998).

13. I. O. Troyanchuk, D. D. Khalyavin, M. Hervieu, *et al.*, Phys. Status Solidi A **169**, R1 (1998).

14. I. O. Troyanchuk, L. S. Lobanovsky, N. V. Kasper, *et al.*, Phys. Rev. B **58**, 14903 (1998).

15. B. Bochu, J. Chenavas, J. C. Joubert, and M. Marezio, J. Solid State Chem. **11** (11), 88 (1974).

16. R. D. Shannon, Acta Crystallogr. A **32**, 751 (1976).

17. B. Bochu, M. N. Deschizeaux, J. C. Joubert, *et al.,* J. Phys. (Paris) **38**, C1-95 (1977).

18. C. Zener, Phys. Rev. **82**, 403 (1951).

19. A. Maignan, B. Raveau, C. Martin, *et al.*, J. Solid State Chem. **144**, 224 (1999).

20. N. Zhang, S. Zhang, W. P. Ding, *et al.*, Solid State Commun. **107**, 417 (1998).

21. A. R. West, *Solid State Chemistry and Its Applications* (Wiley, New York, 1984; Mir, Moscow, 1988).

Translated by A. Chikishev

OPTICS,
QUANTUM ELECTRONICS

Dynamics of Two-Mode Radiation in Optical Waveguides with Strong Mode Coupling

I. O. Zolotovskiĭ and D. I. Sementsov

Ul'yanovsk State University, Ul'yanovsk, 432700 Russia
e-mail: sementsov@ulsu.ru

Abstract—Amplitude equations, as well as the effective dispersion and nonlinearity parameters, which define the dynamics of a wave packet formed by two strongly coupled modes, are derived with allowance for the frequency dependence of the linear mode coupling coefficient. These equations are used to study the onset of the modulation instability of the two-mode wave packet, soliton-like pulses, and compression modes. Unlike single-mode systems, the last two effects in optical waveguides may arise for both a negative and positive dispersion of the waveguide material. © 2003 MAIK "Nauka/Interperiodica".

INTRODUCTION

Optically coupled waveguides (OCWs) with strong linear coupling between modes propagating in different channels and essentially nonlinear optical properties have attracted much interest in recent years [1, 2] because of their considerable promise for laser radiation control devices, such as optical switches, couplers, logical elements, etc. Such OCWs can be built around fiber-optic [3] or planar [4] structures. An example of the planar configuration is, for instance, a structure consisting of two strip waveguides on a single-crystalline $Ga_xAl_{1-x}As$ substrate [5]. With the pulse carrier frequency chosen appropriately, this material exhibits positive Kerr-type nonlinearity, making OCWs capable of supporting soliton-like propagation of optical wave packets. OCWs of this type feature much higher (than in standard quartz fibers) nonlinearity parameters and a high mode-coupling parameter [1]. As is well known, the formation of soliton-like pulses in an optical waveguide depends on its nonlinearity and group velocity dispersion [6, 7]. Since the effective parameters of the OCW nonlinearity and dispersion are to a great extent governed by the excitation conditions, the output radiation can be efficiently controlled by varying the conditions at the input to the optical waveguide.

Of fundamental and applied interest is the modulation instability of radiation in nonlinear OCWs, i.e., time-varying disturbances observed against a background of a sufficiently intense quasi-continuous pumping radiation [6]. Modulation instability transforms a quasi-stationary wave into a periodic sequence of pulses with a period depending on the disturbance frequency. This effect may be used to generate pulses of a given duration and frequency [7].

This paper studies the dynamics of a wave packet in OCWs that provide a strong linear coupling between codirectional waves propagating in the waveguide's channels and possessing a Kerr-type nonlinearity. Effective nonlinearity and dispersion parameters that govern the dynamics of a wave packet in these OCWs are introduced. The onset of modulation instability, as well as the formation and compression of a soliton-like pulse, are studied. The dispersion of the linear mode coupling coefficient is taken into account, because the frequency spectrum of short pulses is wide and its components may have a different mode coupling. This difference is the greatest for planar GaAs or InSb OCWs, where the coupling coefficient may strongly depend on frequency [8, 9].

(1) The propagation of a two-wave optical pulse with allowance for linear coupling dispersion wave detuning, group velocity dispersion, and Kerr-type nonlinearity in a waveguiding medium is described by the following system of equations for the envelopes of two ($j = 1, 2$) interacting waves [10, 11]:

$$
\frac{\partial A_j}{\partial z} + \frac{\xi_j}{v}\frac{\partial A_j}{\partial \tau} - i\frac{d_j}{2}\frac{\partial^2 A_j}{\partial \tau^2} + i\gamma_j|A_j|^2 A_j
$$
$$
= -i\sigma A_{3-j}\exp(-i\xi_j\delta z). \tag{1}
$$

Here, $\tau = t - z/u$ is running time, where $2u = (u_1 + u_2)$ is the group velocity of the wave packet; $u_j = (\partial\beta_j/\partial\omega)^{-1}_{\omega_0}$ is the group velocity; β_j is the propagation constant in a jth channel; ω_0 is the carrier frequency of the wave packet; $d_j = (\partial^2\beta_j/\partial\omega^2)_{\omega_0}$ are the parameters of group velocity dispersion; $v^{-1} = (u_1 - u_2)/2u^2$ is the mode group velocities mismatch; $\delta = \beta_1 - \beta_2$; γ_j are the nonlinearity parameters of the optical waveguide, which define the phase self-modulation of the interacting waves; and $\xi_j = (-1)^j$. The linear coupling parameter σ depends on the overlap of waveguide mode profile functions [2]. The mode coupling dispersion may be

taken into account if the parameter σ is replaced by the operator $\hat{\sigma}_j$ [8, 12]:

$$\hat{\sigma}_j = \sigma\left(1 - \xi_j i\mu \frac{\partial}{\partial\tau}\right), \quad \mu \cong \frac{2}{\omega} + \frac{1}{\sigma}\frac{\partial\sigma}{\partial\omega}. \quad (2)$$

Equations (1) must be solved jointly with the initial conditions for the mode time envelopes A_j. These conditions are specified by the conditions of optical waveguide excitation. The general form of the initial conditions is $A_{20} = \psi A_{10}$, where $A_{i0} = A_i(\tau, 0)$ and the parameter ψ specifies the type of fiber excitation. At $\psi = \pm 1$, the optical waveguide is excited symmetrically or antisymmetrically; at $\psi^{-1} = 0$, one mode is excited.

When solving Eqs. (1) in view of (2), we assume that wave coupling in the wave packet is strong. For this approximation to be valid, the mode interaction length $L_\sigma = |\sigma|^{-1}$ must be much shorter than the dispersion length $L_d = \tau_0^2/|d|$, nonlinearity length $L_n = (\gamma I_0)^{-1}$, and mode spacing $L_v = v\tau_0$. Here, τ_0 and $I_0 = |A_{10}|^2 + |A_{20}|^2$ are, respectively, the duration and intensity of the pulse applied to the optical waveguide (below, by the radiation intensity we mean the power delivered to the waveguide); $2d = d_1 + d_2$; and $2\gamma = \gamma_1 + \gamma_2$. In particular, for coupled strip waveguides fabricated in a GaAs crystal, the coupling parameter was as high as $|\sigma| \cong 5 \times 10^3$ m^{-1} at the wavelength $\lambda = 1.06$ μm, which yields a small mode interaction length $L_\sigma \cong 2 \times 10^{-4}$ m. For the dispersion and nonlinearity parameter values $|d| \cong 10^{-26}$ s^2/m and $\gamma \cong 5 \times 10^3$ (W m)$^{-1}$, which are typical of these materials at $\lambda = 1.06$ μm, the strong-coupling approximation adequately describes the dynamics of the wave packet for input pulses with a duration $\tau_0 \ge 10^{-14}$ s and intensity $I_0 \le 1$ W.

In view of the above, the envelope of the respective mode may be represented as the superposition of two partial pulses:

$$A_1 = a_1(\tau, z)\exp[i(q + \delta/2)z] + a_2(\tau, z)\exp[-i(q - \delta/2)z],$$

$$A_2 = \chi a_1(\tau, z)\exp[i(q - \delta/2)z] \quad (3)$$

$$- \chi^{-1} a_2(\tau, z)\exp[-i(q + \delta/2)z].$$

Here, the amplitudes a_f of the partial pulses are slowly varying functions of z, $q \equiv (\sigma^2 + \delta^2/4)^{1/2}$, and

$$\chi = \frac{(2q + \delta)A_{20} - 2\sigma A_{10}}{(2q - \delta)A_{10} - 2\sigma A_{20}} = -\frac{(2q + \delta)A_{10} + 2\sigma A_{20}}{(2q - \delta)A_{20} + 2\sigma A_{10}} \quad (4)$$

is a parameter depending on the initial conditions of waveguide excitation

When the phase synchronism is exact, $\delta = 0$, $q = |\sigma|$, and $\chi = -1$.

Thus, the pulse formed by two interacting modes is the superposition of partial pulses whose amplitudes

satisfy the equations

$$\frac{\partial a_f}{\partial z} - \frac{\xi_f \delta}{2qv}\frac{\partial a_f}{\partial\tau} - \frac{iD_f}{2}\frac{\partial^2 a_f}{\partial\tau^2} \quad (5)$$

$$+ i(G_{sf}|a_f|^2 + G_{cf}|a_{3-f}|^2)a_f = 0$$

in accordance with (1) and (3). Here, $f = 1$ or 2 is the pulse index and

$$G_{sf} = [[q + \xi_f\delta/2]\gamma_{s1} + [q - \xi_f\delta/2]\chi^{-2\xi_f}\gamma_{s2}]/2q,$$

$$G_{cf} = [[2q + \xi_f\delta]\gamma_{s1} \quad (6a)$$

$$+ [q(\chi^{2\xi_f} + 1) - \xi_f\delta(\chi^{2\xi_f} - 1)/2]\gamma_{s2}]/2q,$$

$$D_f = \frac{d_1 + d_2}{2} + \frac{\xi_f}{v^2 q}\left(1 - \frac{\delta v^2(d_2 - d_1)}{2} + (v\sigma\mu)^2\right) \quad (6b)$$

are the effective parameters of phase self-modulation, cross modulation, and dispersion.

By virtue of (3), the initial conditions for the partial pulse amplitudes take the form

$$a_f(\tau, 0) \equiv a_{f0} = \frac{1}{2}\left[A_{10} + \xi_f\left(\frac{\sigma}{2q}A_{10} + \frac{\sigma}{q}A_{20}\right)\right]. \quad (7)$$

It is important to note that the system of equations (5) involves the parameters of induced cross-modulation interaction, G_{cf}, which are absent in the initial system of equations (1). The cross-modulation interaction arises when partial pulses into which a wave packet applied to the OCW breaks interact between each other in either of the waveguides. Of significance is also the strong dependence of the effective parameters on the initial excitation conditions, phase mismatch, and mode coupling coefficient dispersion. For instance, when the waves are in phase synchronism ($\delta = 0$), the contribution of the mode interaction dispersion to the effective dispersion is, according to (2), $d_\sigma = \pm\mu^2\sigma \cong \pm 4\pi^2\sigma/\omega_0^2$. For GaAs OCWs, this parameter usually lies in the interval $|d_\sigma| = 0.1$–10×10^{-26} s^2/m (at the carrier frequency of the source $\omega_0 \cong 1.8 \times 10^{16}$ s^{-1}). Since the parameter $|d|$, which is associated with the material dispersion, also varies within this interval, the dispersion of the linear coupling should necessarily be taken into account when the pulse dynamics is analyzed.

(2) Consider conditions for modulation instability in the OCWs under study. If the wave packet applied is long enough, so that the pumping wave can be described in terms of the quasi-monochromatic approximation (i.e., the dispersion terms are negligible, which seems to be correct for pulses with an initial duration of $\tau_0 > 10^{-9}$ s), the solution to Eqs. (5) for the partial pulse amplitudes may be represented as

$$a_f(z, \tau) = [a_{f0} + \varphi_f(z, \tau)]$$

$$\times \exp[-iR(G_{sf}a_{f0}^2 + G_{cf}a_{3-f0}^2)], \quad (8)$$

where φ_f is the complex disturbance amplitude and $a_{f0} \gg |\varphi_f|$.

Next, we substitute (8) into Eqs. (5) and linearize them in small disturbances φ_f to obtain a system of equations for the disturbances:

$$\frac{\partial \varphi_f}{\partial z} - \xi_f \frac{\delta}{2qv} \frac{\partial \varphi_f}{\partial \tau} - \frac{iD_f}{2} \frac{\partial^2 \varphi_f}{\partial \tau^2} + iG_{sf}a_{f0}^2(\varphi_f + \varphi_f^*)$$
$$+ iG_{cf}a_{10}a_{20}(\varphi_{3-f} + \varphi_{3-f}^*) = 0. \quad (9)$$

For harmonic disturbances $\varphi_f(z, \tau)$, a solution to this equation will be sought in the standard form

$$\varphi_f = b_{f1}\cos(kz - \Omega\tau) + ib_{f2}\sin(kz - \Omega\tau), \quad (10)$$

where k and Ω are the wavenumber and frequency of the disturbance ($\Omega = \omega_0 - \omega_v$, where ω_v is the frequency of a disturbing signal wave or spontaneous noise disturbance).

Substituting (10) into (9) yields a system of two equations for the amplitudes b_{fi}. A solution to this system is the dispersion relation

$$\left[\left(k - \frac{\delta\Omega}{2qv}\right)^2 - K_1\right]\left[\left(k + \frac{\delta\Omega}{2qv}\right)^2 - K_2\right] = F,$$
$$K_f = 0.25D_f\Omega^2(D_f\Omega^2 + 4G_{sf}a_{f0}^2), \quad (11)$$
$$F = D_1D_2G_{c1}G_{c2}\Omega^4 a_{10}^2 a_{20}^2.$$

First, consider the important particular case $F = 0$. This situation may occur when the effective dispersion or the amplitude of one of the partial pulses is zero. Let, for example, $D_2 = 0$. Hence, $K_2 = 0$. In this case, Eq. (11) has the solutions

$$k_1 = \frac{\delta\Omega}{2qv} \pm \sqrt{K_1}, \quad k_2 = -\frac{\delta\Omega}{2qv}. \quad (12)$$

If the parameter ranges are such that $K_1 < 0$, modulation instability is seen to be caused by the solution with the wave number k_1. Since G_{s1} is positive, modulation instability develops if the parameters meet the condition

$$-4G_{s1}a_{10}^2 < \Omega^2 D_1 < 0, \quad (13)$$

which can be realized by appropriately choosing the disturbance frequency or the intensity of the pulse applied. In this case, modulation instability takes place at an anomalous effective dispersion, i.e., at $D_1 < 0$. It should be emphasized that the material dispersion parameter may be positive ($d > 0$). This situation is fundamentally different from the conventional single-mode case, where modulation instability exists only when the material dispersion parameter is negative. Similar modulation instability conditions for one of the branches of the solution to the dispersion relation take place when the amplitude of one partial pulse is zero. For example, with $a_{20} = 0$, solutions of Eq. (11) are

identical to those of (12). As in the case of $D_2 = 0$, here modulation instability arises when condition (13) is met. The increment of the propagating disturbance in the above situations is given by

$$g = 2\mathrm{Im}k_1 = |D_1\Omega|\sqrt{\Omega_s^2 - \Omega^2}, \quad (14)$$

where $\Omega_s^2 = 4G_{s1}a_{10}^2/|D_1|$. The increment reaches a maximum,

$$g_m = |D_1|\Omega_s^2/2 = 2G_{s1}a_{10}^2,$$

at a disturbance frequency $\Omega = \Omega_s/\sqrt{2}$.

Figure 1 plots the increment versus disturbance frequency when one partial pulse propagates in the system. The curves $g(\Omega)$ were obtained at $\sigma = 5 \times 10^3$ m^{-1}, $\gamma_{s1} = \gamma_{s2} = 5 \times 10^3$ (W m)$^{-1}$, $d_1 = -0.975 \times 10^{-26}$ s^2/m, $d_2 = -1.025 \times 10^{-26}$ s^2/m, $\mu = 4 \times 10^{-16}$ s, $v^{-1} = 10^{-12}$ s/m, and various phase mismatches. According to (7), to realize the single-pulse regime with $a_{20} = 0$ at the above values of $\delta = 10$, we must respectively take $\psi = (1) - 1.001$, (2) -1.01, and (3) -1.1; i.e., the excitation of the optical waveguide must be almost antisymmetric. As can be seen from Fig. 1, the increase in the phase mismatch extends the frequency range of modulation instability and raises the disturbance increment.

Let us focus on the practically important case of exact phase synchronism ($\delta \cong 0$). Dispersion relation (11) is then reduced to a biquadratic equation whose solution is

$$k_\pm^2 = \frac{K_1 + K_2}{2} \pm \sqrt{\frac{(K_1 - K_2)^2}{4} + F}. \quad (15)$$

Here, there are several scenarios of modulation instability development, of which we distinguish the main two. The first one is realized if

$$k_+^2 k_-^2 = K_1 K_2 - F < 0. \quad (16)$$

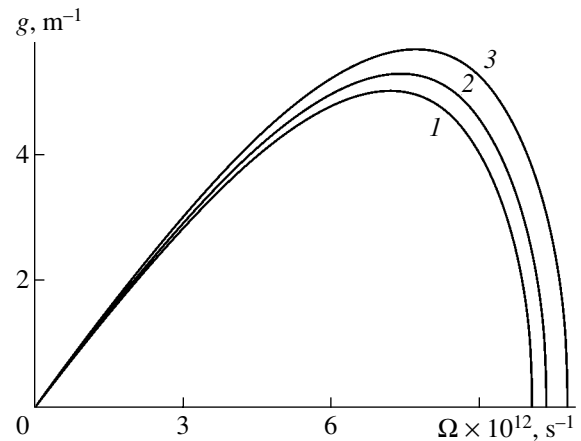In this case, modulation instability is observed only

Fig. 1. Modulation instability increment versus disturbance frequency. $\delta = (1)$ 10, (2) 100, and (3) 1000 m^{-1}.

in one branch of the solution to Eq. (11), specifically, for the disturbance with the wavenumber k_-. The increment of the propagating disturbance is then given by the expression

$$g_- = 2\mathrm{Im}(k_-)$$
$$= \sqrt{2}[\sqrt{(K_1 - K_2)^2 + 4F} - K_1 - K_2]^{1/2}. \quad (17)$$

Condition (16) specifies the disturbance frequency range in which modulation instability occurs, $|\Omega| < \Omega_{s1}$, where

$$\Omega_{s1} = 2$$
$$\times (\sqrt{(G_{s1}a_{10}^2/D_1 - G_{s2}a_{20}^2/D_2)^2 + 4a_{10}^2 a_{20}^2 G_{c1}G_{c2}/D_1 D_2} \quad (18)$$
$$- G_{s1}a_{10}^2/D_1 - G_{s2}a_{20}^2/D_2)^{1/2}.$$

As follows from (18), modulation instability here may develop at any sign of the effective partial pulse dispersion parameters, unlike the previous case. It is necessary to take into account that the relationship $G_{cf} \geq G_{sf}$ is always valid for OCWs with strong linear coupling. When the effective dispersions of both partial pulses are normal ($D_1 > 0$ and $D_2 > 0$), the situation differs from the single-pulse propagation most significantly. However, the maximum values of the increment g_- and disturbance frequency Ω_{s1} up to which modulation instability is observed turn out to be lower than those when $D_1 < 0$ and $D_2 < 0$.

The situation when two branches of solution (14) are responsible for modulation instability is fundamentally different from the one observed in single-pulse systems. In this case, both roots k_+ and k_- are purely imaginary. This situation is realized when the conditions

$$K_1 + K_2 < 0, \quad K_1 K_2 - F > 0 \quad (19)$$

are fulfilled in the frequency range $\Omega_{s1} < \Omega < \Omega_c$, where

$$\Omega_c = 2[-(D_1 G_{s1}a_{10}^2 + D_2 G_{s2}a_{20}^2)(D_1^2 + D_2^2)^{-1}]^{1/2}. \quad (20)$$

Conditions (19) are satisfied when the effective dispersions of the partial pulses are equal but opposite in sign. The increment is then given by the expression

$$g_\pm = 2\mathrm{Im}(k_\pm)$$
$$= \sqrt{2}[|K_1 + K_2| \mp \sqrt{(K_1 - K_2)^2 + 4F}]^{1/2}. \quad (21)$$

As follows from the above expressions, modulation instability in OCWs with a strong mode coupling may be controlled by varying the effective parameters of partial pulse nonlinearity and dispersion.

(3) Another problem of practical interest is the feasibility of soliton-like pulses and compression modes in the optical waveguides under study. Of most interest is obtaining analytical solutions to Eqs. (5) and their experimental implementation in the degenerate situation where the whole wave packet may be described by only one of the partial pulses. For this degenerate situation, one can find the optimal dispersion parameters of the pulse propagating in the optical waveguide. If the modes are in phase synchronism ($\delta = 0$), degeneration takes place when the excitation of the waveguide is symmetric ($\psi = 1$) or asymmetric ($\psi = -1$) with $a_{10} = 0$ and $a_{20} \neq 0$ or $a_{20} = 0$ and $a_{10} \neq 0$, respectively. In the case of a phase mismatch ($\delta \neq 0$), a degenerate situation is also possible for asymmetric excitation, as follows from (7). In particular, if $\psi_f = (2\xi_f q - \delta)/2\sigma$, the partial pulse amplitudes are $a_f \neq 0$ and $a_{3-f} = 0$. The amplitude of the corresponding partial pulse is now zero not only at the initial moment but also throughout the pulse propagation. In the above cases, where the amplitude of one of the partial pulses vanishes, the system of equations (5) degenerates into one nonlinear Schrödinger equation

$$\frac{\partial a_f}{\partial z} - \frac{iD_f}{2}\frac{\partial^2 a_f}{\partial \tau_f^2} + iG_{sf}|a_f|^2 a_f = 0, \quad (22)$$

where $\tau_f = t - z/u_f$ is time in the running coordinate system related to the corresponding partial pulse, $u_f = u/(1 - \Delta_f u)$ is its group velocity, and $\Delta_f = \xi_f \delta/2qv$.

Equation (22) describes the pulse dynamics in a cubically nonlinear medium with an effective dispersion D_f and effective nonlinearity G_{sf}. The formation dynamics of one-soliton and multisoliton pulses, self-compression effects, etc., for similar equations have been studied in detail [6, 13]. It is important that the pulse dynamics is governed by the effective parameters of dispersion and nonlinearity, which here depend not only on the material properties but also on the initial conditions of OCW excitation.

If the system of equations (5) cannot be reduced to the degenerate case, its exact analytical solution is impossible to find. In this case, it can be solved by the variational method, which has been successfully applied to many problems of nonlinear optics [14, 15]. If the phase mismatch is small ($|\delta/\sigma| < 1$), the duration of the pulse applied is expressed as

$$\left(\frac{d\tau_p}{dz}\right)^2 = f_0 - 4D(W_1 + W_2)G/\pi^2\tau_p - 4D^2/\pi^2\tau_p^2. \quad (23)$$

Here, $f_0 = 4D(W_1 + W_2)G/\pi^2\tau_0 + 4D^2/\pi^2\tau_0^2$,

$$D = (W_1 D_1 + W_2 D_2)/(W_1 + W_2),$$
$$G = (G_{s1}W_1^2 + G_{s2}W_2^2 + (G_{c1} + G_{c2})W_1 W_2)/(W_1 + W_2) \quad (24)$$

are the effective dispersion and nonlinearity parameters of the wave packet, and

$$W_f = I_{f0}\tau_0 = W_0[1 + \xi_f(\delta + 2\sigma)/2q]^2/4(1 + \psi^2) \quad (25)$$

is the partial pulse energy, where $W_0 = I_0\tau_0$ is the energy of the radiation applied to the optical waveguide.
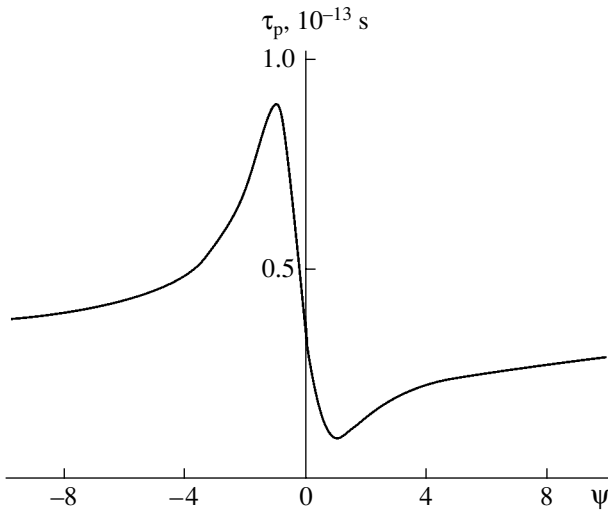
Fig. 2. Pulse duration versus the optical waveguide excitation parameter.

In the case of exact phase synchronism ($\delta = 0$), the expressions for the effective nonlinearity and dispersion parameters take the form

$$G = \frac{\gamma_{s1} + \gamma_{s2}}{2}\left[1 + \frac{1}{2}\left(\frac{1 - \psi^2}{1 + \psi^2}\right)^2\right],$$

$$D = \frac{d_1 + d_2}{2} + \left(\frac{2\psi}{1 + \psi^2}\right)\frac{1 + (v\sigma\mu)^2}{v^2\sigma}. \quad (26)$$

Equation (23) implies that the problem is reduced to the well-known Kepler equation [16]. Omitting the solution and analysis of this equation, we note that, when $D < 0$, the condition $f_0 = 0$ makes it possible to determine the boundary value of the energy applied to the fiber, W_0, above which a quasi-soliton pulse, i.e., a nonspreading solitary wave, forms. In this case, Eq. (23) describes a pulse whose duration varies periodically about $\langle \tau_p \rangle = 2|D|/(W_1 + W_2)G$. If $\tau_0(W_1 + W_2)G/2|D| > 1$, the pulse compression mode is realized. For example, by applying a pulse of duration $\tau_0 = 4 \times 10^{-13}$ s and power $I_0 = 10^{-2}$ W to an optical waveguide with $G \cong 5 \times 10^3$ (W m)$^{-1}$ and $D \cong -10^{-26}$ s^2/m, a soliton-like pulse of extremely short duration $t_s \cong 10^{-14}$ s can be generated over a length $L_s \cong \sqrt{\tau_0^2/|D| G I_0} \cong 0.15$ m along the optical waveguide. Figure 2 plots the pulse duration τ_p versus the waveguide excitation parameter ψ after the pulse has traveled the length L_s. The curve was constructed at $\tau_0 = 4 \times 10^{-13}$ s, $I_0 = 10^{-2}$ W, and $\delta =$

0. The values of the dispersion and nonlinearity parameters were the same as above. For $\psi \cong 1$ (symmetric excitation), the pulse duration is minimum, $\tau_p \cong 10^{-14}$ s. The pulse compression in this case reaches $\tau_0/\tau_p \cong 40$.

The above analysis shows that the dependence of the effective dispersion and nonlinearity parameters on the detuning, mode coupling, and type of OCW excitation makes it possible to effectively control modulation instability, the formation of soliton-like pulses, the degree of compression, and other parameters that specify the dynamics of a wave packet propagating in OCWs. Similar results are expected for OCWs using light fibers and crystals, where a small spacing between the guiding channels provides a strong wave coupling.

REFERENCES

1. A. A. Maĭer, Usp. Fiz. Nauk **165**, 1037 (1995) [Phys. Usp. **38**, 991 (1995)].
2. F. Kh. Abdullaev, Zh. Tekh. Fiz. **68** (6), 1 (1998) [Tech. Phys. **43**, 615 (1998)].
3. A. A. Maĭer, Kvantovaya Elektron. (Moscow) **2**, 1490 (1982).
4. S. M. Jensen, IEEE J. Quantum Electron. **18**, 1580 (1982).
5. D. B. Mortimor and J. M. Arkwright, Appl. Opt. **29**, 1814 (1990).
6. G. P. Agraval, *Nonlinear Fiber Optics* (Academic, New York, 1989; Mir, Moscow, 1996).
7. S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *The Optics of Femtosecond Laser Pulses* (Nauka, Moscow, 1988), p. 310.
8. A. A. Maĭer and S. G. Karataev, Kvantovaya Élektron. (Moscow) **23** (1), 43 (1996).
9. Y. J. Chen and G. M. Carter, Appl. Phys. Lett. **41**, 307 (1984).
10. I. O. Zolotovskiĭ and D. I. Sementsov, Opt. Spektrosk. **86**, 737 (1999) [Opt. Spectrosc. **86**, 737 (1999)].
11. I. O. Zolotovskiĭ and D. I. Sementsov, Kvantovaya Élektron. (Moscow) **27**, 273 (1999).
12. I. O. Zolotovskiĭ and D. I. Sementsov, Pis'ma Zh. Tekh. Fiz. **28** (24), 30 (2002) [Tech. Phys. Lett. **28**, 1033 (2002)].
13. I. V. Dzedolik and A. I. Dzedolik, Zh. Tekh. Fiz. **72** (6), 61 (2002) [Tech. Phys. **47**, 713 (2002)].
14. D. Anderson, Phys. Rev. A **27**, 3135 (1983).
15. A. I. Maĭmistov, Kvantovaya Élektron. (Moscow) **18**, 758 (1991).
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics* (Nauka, Moscow, 1988; Pergamon, New York, 1988).

Translated by A. Khzmalyan

<center>═══════ **RADIOPHYSICS** ═══════</center>

Transition to Fully Developed Chaos in a System of Two Unidirectionally Coupled Backward-Wave Oscillators

N. M. Ryskin and V. N. Titov

Saratov State University, Moskovskaya ul. 155, Saratov, 410012 Russia
e-mail: RyskinNM@info.sgu.ru
Received February 20, 2003

Abstract—The chaotic dynamics of a system of two unidirectionally coupled backward-wave oscillators (BWOs) is studied in the case when a signal from the driving BWO in (periodic or chaotic) self-modulation mode is applied to the driven oscillator, which exhibits strong periodic self-modulation in the autonomous case. The oscillation evolution with the amount of coupling is traced. The use of a chain of coupled BWOs is shown to significantly reduce the threshold of transition to the regime of wide-band chaotic oscillations with a uniform continuous spectrum (so-called fully developed chaos), which is of interest for applications. © 2003 MAIK "Nauka/Interperiodica".

High-power sources of microwave chaotic oscillations are promising for a wide range of applications, such as radar detection, plasma heating in systems of controlled thermonuclear fusion, advanced communications based on dynamic chaos, etc. Backward-wave oscillators are among the most extensively studied complex-dynamics vacuum electronic devices: their capability to generate chaotic oscillations was predicted and experimentally confirmed as early as in the late 1970s [1, 2]. Recent studies of BWO operation [3–7] showed that, as the beam current increases, regular and chaotic oscillation regimes alternate in a complex manner until the system comes to exhibit highly irregular wide-band chaotic oscillations with a fairly uniform continuous spectrum. This mode has received the name fully developed chaos. Evidently, it is the most favorable regime for the applications mentioned above. However, fully developed chaos arises when the electron beam current is considerably in excess (by more than 30 times) of the starting value, which is hard to provide in reality. Experiments with nonrelativistic [2, 8] and relativistic [9–11] BWOs are usually carried out on specially constructed setups with an extended phase length. In this case, additional difficulties associated, for example, with beam focusing may emerge.

In this study, we will show that two coupled oscillators may reduce the developed chaos threshold. The nonlinear dynamics of such a system is numerically studied by using the well-known equations from the nonstationary nonlinear theory of BWO (see, e.g., [1, 3–7]):

$$\frac{\partial^2 \theta_{1,2}}{\partial \xi^2} = -L_{1,2}^2 \mathrm{Re}[F_{1,2} \exp(i\theta_{1,2})], \qquad (1)$$

$$\frac{\partial F_{1,2}}{\partial \tau} - \frac{\partial F_{1,2}}{\partial \xi} = -\frac{L_{1,2}}{\pi} \int_0^{2\pi} \exp(-i\theta_{1,2})\,d\theta_0. \qquad (2)$$

Equations (1) describe the motion of electrons in the field of an electromagnetic wave, while Eqs. (2) represent the nonstationary excitation of the slow-wave structure by a slowly varying current. The subscripts indicate the serial number of an oscillator in the chain. In Eqs. (1) and (2), $\theta_{1,2}$ are the electron phases with respect to the wave; θ_0 are the initial phases; $F_{1,2}$ are the slowly varying dimensionless amplitudes; and ξ and τ are the dimensionless coordinate and time, respectively. Note that the equations of motion are written under the assumption of a small variation of electron energy during the interaction (see, e. g., [1, 6]) and apply to both the relativistic and nonrelativistic cases. The dynamics of a partial oscillator is governed by the sole bifurcation parameter $L = 2\pi CN$, where C is the Pierce gain parameter and N is the phase length of the system. If one oscillator is taken to drive the other, the boundary conditions for Eqs. (1) and (2) can be set as

$$\theta_{1,2}|_{\xi=0} = \theta_0 \in [0; 2\pi], \quad \left.\frac{\partial \theta_{1,2}}{\partial \xi}\right|_{\xi=0} = 0,$$

$$F_1(\xi=1) = 0, \quad F_2(\xi=1) = RF_1(\xi=0), \qquad (3)$$

where R is the coupling parameter, which may be considered real without loss of generality.

Let us recall the basic results of the investigation into the complex dynamics of an autonomous oscillator [3, 4]. The self-excitation of oscillations starts at $L \approx 1.98$. Stationary single-frequency oscillation modes exist within the range $1.98 < L < 2.9$. At $L \approx 2.9$, the stationary regime becomes unstable and gives way to periodic self-modulation. As the parameter L increases to

≈4.02, the periodic self-modulation persists with both the fundamental and self-modulation frequencies remaining almost unchanged. Then, in the narrow range of the bifurcation parameter $4.02 < L < 4.11$, chaos is established via a cascade of period doubling bifurcations. After the chaos threshold has been exceeded (at $4.150 < L < 4.3325$), numerous windows of periodicity appear, with their periods successively increasing by unity starting from three (the range of period add-on). Windows with periods up to 11 were reliably detected in [3, 4]. At $L > 4.333$, the transition from the chaotic regime to regular self-modulation is observed and this transition occurs via intermittency. In this case, the strange attractor turns into a metastable chaotic set. During the transient, the phase trajectory first lies in the vicinity of this set (the waveform resembles a chaotic one) and then is attracted to a stable limit cycle. Note that the route via intermittency is accompanied by a sharp shift of the fundamental frequencies in the spectrum. As was shown in [6], the reverse transition to the regular regime causes a profound transformation of the space–time structures that form in the electron beam at the stage of intense beam overbunching. Next comes the range $4.625 < L < 4.750$ with quasiperiodic self-modulation at two incommensurable frequencies. Narrow windows of synchronization are observed in this range when the self-modulation fundamental frequency ratio is rational. This situation corresponds to the on-torus formation of resonant cycles. At $L > 4.75$, regimes are periodic again. Eventually, the cascade of bifurcations brings the system, at $L ≈ 6.05$, to the state of fully developed chaos, which is characterized by a uniform continuous spectrum and the absence of any large-scale structure on the projection of the phase portrait. This transition also takes place via intermittency.

Let us next consider a chain of two unidirectionally coupled BWOs. The construction of a complete dynamic model for such a system is a great challenge and goes beyond the scope of this study. Here, we will single out various situations depending on the operating mode of either oscillator. For example, one may consider the action of a periodic-mode oscillator on an oscillator operating in the chaotic mode and vice versa, interaction between two chaotic-signal sources, etc. Note that, if the driving oscillator operates in the steady single-frequency mode, we face the well-known problem of synchronization by an external harmonic signal. The problem of synchronization of a periodic oscillator has been studied in most detail; in general, this problem is qualitatively similar to the classical problem of the vibration theory when a harmonic force acts on a self-oscillating system [13, 14]. The effect of a harmonic signal on a chaotic BWO was studied in [8], where chaos was found to be suppressed with growing external signal. This phenomena calls for further investigation; however, it is safe to predict the presence of all the effects typical of finite-dimensional chaotic systems that are synchronized by an external harmonic action,

such as the locking or suppression of the fundamental frequency of chaotic oscillations, etc. [15].

Since the primary goal of this study is to see whether the developed chaos threshold may be reduced, we will first consider the situation where both oscillators operate in the periodic self-modulation mode. The evolution of the oscillatory regimes with the amount of coupling seems to be the following. As R grows, oscillations become chaotic basically through the decay of the quasiperiodic motion. Figure 1 presents typical phase portraits and output signal spectra for (a) the first and (b–f) second oscillators. The two-dimensional projection of the phase portrait is constructed for $F_{out} = |F_{1, 2}(\xi = 0)|$ by the method of delays (Packard–Takens method) [15]. The values of the bifurcation parameters $L_1 = 4.0$ and $L_2 = 4.5$ are chosen so as to ensure autonomous operation in the strong periodic self-modulation mode (Figs. 1a, 1b) but away from the developed chaos threshold (the current exceeds the initial value eight and eleven times, respectively). In the first case, we are near the Feigenbaum chaos threshold; the second case corresponds to the situation after the chaos–order transition via intermittency. It is seen that the phase portraits and spectra in these two cases are qualitatively different.

With an increase in the degree of coupling, the oscillations first become quasiperiodic (Fig. 1c) and then chaotic (Fig. 1a) even if coupling is rather weak. The discrete spectral components are distinctly seen against the background of the low noise pedestal. Next, the oscillations become still more irregular, the discrete components decrease and diffuse, while the noise pedestal grows (Fig. 1e). However, at sufficiently high R, new discrete peaks appear, this time, at the frequencies contained in the spectrum of the driving signal (Fig. 1f). Moreover, under certain conditions, self-modulation may again become periodic (for parameter values other than those shown in Fig. 1). Thus, there is an optimal value of R corresponding to the most noisy uniform spectrum.

Note that, in the presence of coupling, both the average output power and the efficiency are higher than those in the case of autonomous operation. This readily follows from the following considerations [12]. In an autonomous oscillator, the distributions of the field and current over the region of interaction are in "antiphase": near the collector, the beam is well bunched but the field is small; therefore, energy exchange is inefficient. Coupling "corrects" the field structure of the output (driven) tube, making it more uniform and providing more favorable conditions for energy removal from electrons.

If the oscillators are identical ($L_1 = L_2$) and operate near the Feigenbaum chaos threshold, the coupling-induced transition to chaos also follows the Feigenbaum scenario. However, this variant is encountered more rarely than the above-mentioned quasiperiodic route, which is always observed when the oscillators
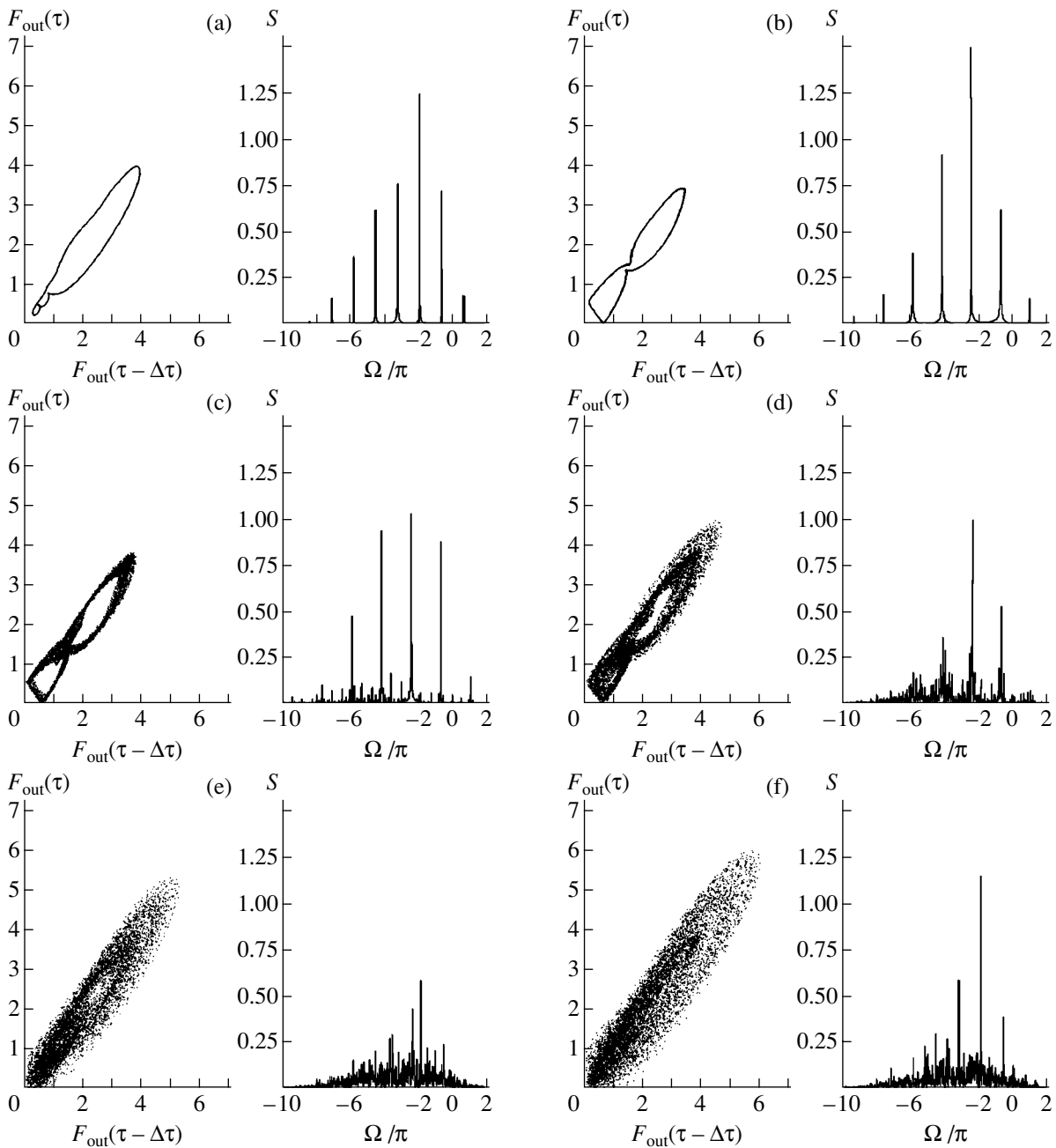
Fig. 1. Phase portraits and spectra of the (a) applied and (b–f) output signals when both oscillators operate in the periodic self-modulation mode ($L_1 = 4.0$, $L_2 = 4.5$) at $R =$ (b) 0, (c) 0.02, (d) 0.05, (e) 0.2, and (f) 0.5.

are nonidentical ($L_1 \neq L_2$) or the frequencies are detuned.[1] Generally, detuning favors the transition to chaos.

With the first oscillator operating in the regime of weakly developed chaos, the behavior of the system is only slightly different from that discussed previously. There is no point in talking of a scenario of transition to chaos in this case; indeed, a chaotic signal, small as it may be, when applied to the second tube, gives rise to a chaotic component in the output signal. However, the evolution of the spectra with the degree of coupling proceeds in the same manner: first, the discrete components diffuse and the noise pedestal grows; then, at sufficiently high R, the fundamental frequencies of the applied signal spectrum start showing up.

---

[1] In Eqs. (1) and (2), $F_{1, 2}$ stands for the amplitudes of waves whose carrier frequencies are taken to be equal to the wave–beam synchronism frequencies. If these frequencies are unequal for both tubes, one should replace R in Eq. (3) by $R\exp(i\Delta\tau)$, where Δ is the frequency detuning.
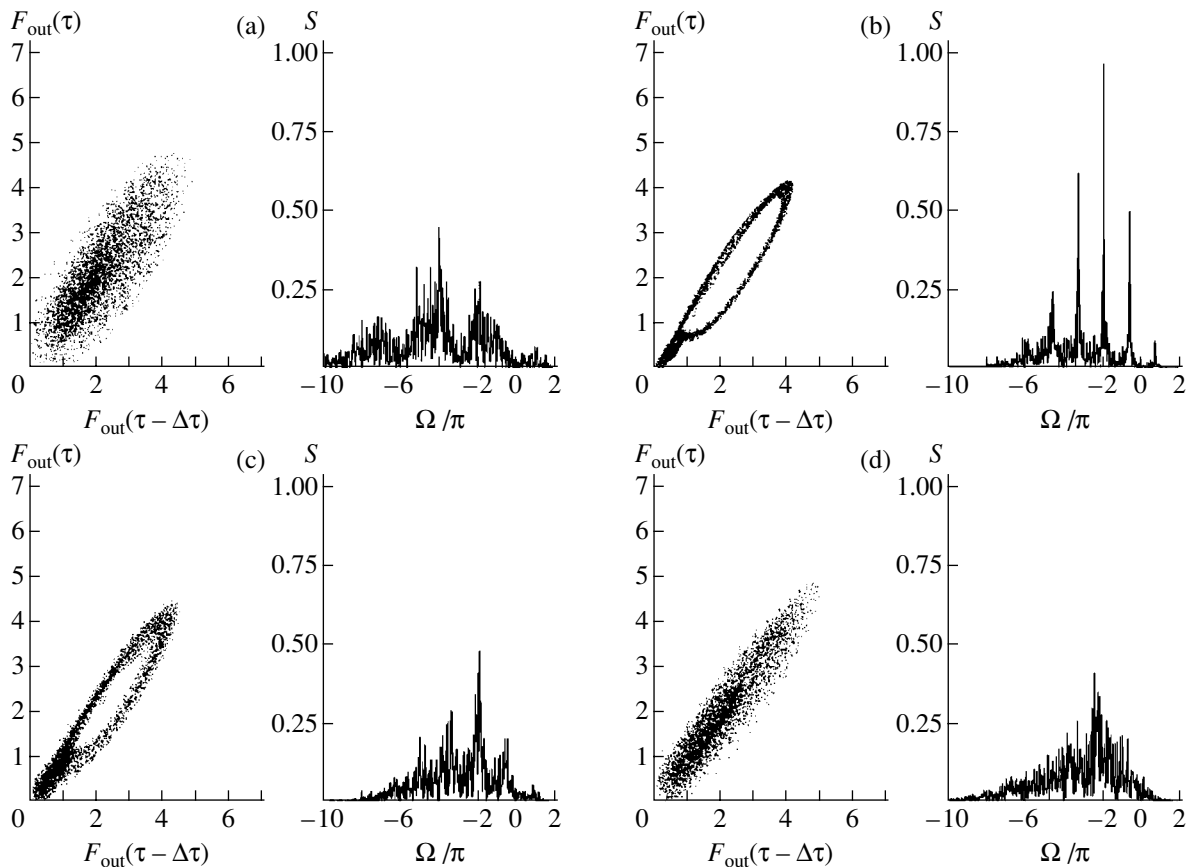
Fig. 2. Phase portraits and spectra of the (a) applied chaotic and (b–d) output signals. The driven oscillator operates in the periodic self-modulation mode ($L_1 = 6.5$, $L_2 = 4.0$) at $R =$ (b) 0.03, (c) 0.1, and (d) 0.3. The data for the output signal obtained in the autonomous mode is given in Fig. 1a.

One more interesting case is that when a wide-band chaotic low-amplitude signal is applied to a high-power oscillator operating in the single-mode or periodic self-modulation regime. In some sense, this situation is similar to the classical problem of stabilization of a high-power oscillator by a weak external signal [13, 14]. However, we are dealing here with the reverse problem, which consists in initiating the transition of a powerful generator (for which the increase in the current to a desired level is for some reason impossible) to fully developed chaos by applying a small chaotic signal with a broad spectrum. Note that earlier a similar idea was put forward in [16], where the noise or regular multifrequency modulation of the electron beam velocity was proposed.

Let us discuss the effect of the fully developed chaotic signal from the first oscillator on the other oscillator working in the periodic self-modulation mode. Typical results are presented in Fig. 2 for $L_1 = 6.5$ (or $I/I_{st} \approx$ 35) and $L_2 = 4.0$ (or $I/I_{st} \approx 8$). The phase portraits and spectra of the driving and output signals are plotted for different values of R (the phase portrait and spectrum of the output signal obtained in the autonomous mode are given in Fig. 1a). Evidently, it is reasonable to consider

small values of R, since the power of the first oscillator is assumed to be low. Under the assumption that the entire output of the first oscillator is applied to the second tube, the power ratio equals $20 \log R$, which gives approximately –30, –20, and –10 dB for Figs. 2b–2c, respectively. Figure 2 clearly demonstrates that, with an increase in R, the discrete components in the spectrum decay rapidly and the noise pedestal grows. At $R = 0.3$, no distinct structure is present in the phase portrait and the spectrum of the output signal turns out to be even more uniform than that of the applied signal. Such behavior is typical when the second (output) oscillator operates in the periodic self-modulation mode. Yet it is reasonable to consider the case of sufficiently strong self-modulation, since a decrease in L_2 leads one to increase R (i. e., the power applied).

Thus, we considered the complex dynamics of a set of two unidirectionally coupled BWOs. The transition to chaos with growing degree of coupling was studied for the case when autonomous oscillators operate in the periodic self-oscillation mode. It is shown that the quasiperiodic route to chaos is the basic scenario of the process. By using a chain of coupled oscillators, one may appreciably reduce the threshold of transition to

fully developed chaos and obtain chaotic oscillations with a uniform continuous spectrum, the ratio of the current to its starting value being much lower than in the autonomous case. However, there exist optimal coupling values at which the output signal is the most irregular. At a higher degree of coupling, the oscillations may again become periodic, because discrete components in the applied signal spectrum start to prevail. The application of a weak wide-band chaotic signal is shown to be another possible way of transition to fully developed chaos in BWOs. The results obtained are of interest for applications using sources of microwave chaotic oscillations. It should be noted that an exhaustive theoretical description of the complex dynamics of the system discussed is a great challenge and calls for further experimental and theoretical research.

ACKNOWLEDGMENTS

REFERENCES

1. N. S. Ginzburg, S. P. Kuznetsov, and T. N. Fedoseeva, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **21**, 1037 (1978).

2. B. P. Bezruchko, S. P. Kuznetsov, and D. I. Trubetskov, Pis'ma Zh. Éksp. Teor. Fiz. **29**, 180 (1979) [JETP Lett. **29**, 162 (1979)].

3. N. M. Ryskin and V. N. Titov, Izv. Vyssh. Uchebn. Zaved. Prikl. Nelineinaya Din. **6** (1), 75 (1998).

4. N. M. Ryskin, V. N. Titov, and D. I. Trubetskov, Dokl. Akad. Nauk **358**, 620 (1998) [Dokl. Phys. **43**, 90 (1998)].

5. N. M. Ryskin and V. N. Titov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **42**, 566 (1999).

6. N. M. Ryskin and V. N. Titov, J. Commun. Technol. Electron., Suppl. 1, **45** S46 (2000).

7. N. M. Ryskin and V. N. Titov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **44**, 860 (2001).

8. B. P. Bezruchko, L. V. Bulgakova, S. P. Kuznetsov, and D. I. Trubetskov, in *Proceedings of the 5th Winter Workshop for Engineers "Lectures on Microwave Electronics and Radiophysics"* (Saratov. Gos. Univ., Saratov, 1980), Vol. 5, pp. 25–77.

9. N. S. Ginzburg, N. I. Zaĭtsev, E. V. Ilyakov, *et al.*, Pis'ma Zh. Tekh. Fiz. **24** (20), 66 (1998) [Tech. Phys. Lett. **24**, 816 (1998)].

10. N. S. Ginzburg, N. I. Zaĭtsev, E. V. Ilyakov, *et al.*, Izv. Vyssh. Uchebn. Zaved. Prikl. Nelineinaya Din. **7** (5), 60 (1999).

11. N. S. Ginzburg, N. I. Zaĭtsev, E. V. Ilyakov, *et al.*, Zh. Tekh. Fiz. **71** (11), 73 (2001) [Tech. Phys. **46**, 1420 (2001)].

12. A. M. Kats, E. N. Il'ina, and I. A. Man'kin, *Nonlinear Phenomena in O-Type Microwave Devices with Long-Term Interaction* (Sov. Radio, Moscow, 1975).

13. M. I. Rabinovich and D. I. Trubetskov, *Oscillations and Waves in Linear and Nonlinear Systems* (Nauka, Moscow, 1984; Kluwer, Dordrecht, 1989).

14. A. P. Kuznetsov, S. P. Kuznetsov, and N. M. Ryskin, *Nonlinear Oscillations* (Fizmatlit, Moscow, 2002).

15. V. S. Anishchenko, T. E. Vadivasova, and V. V. Astakhov, *Nonlinear Dynamics of Chaotic and Stochastic Systems* (Saratov. Gos. Univ., Saratov, 1999).

16. V. L. Vaks, N. S. Ginzburg, A. S. Sergeev, *et al.*, Radiotekh. Élektron. (Moscow) **39**, 957 (1994).

Translated by A. Sidorova

===================== **RADIOPHYSICS** =====================

Waves in Multilayer Ferrite–Dielectric Transversely Magnetized Waveguides with an Electrically Controlled Slowing Factor

K. V. Guzenko and É. F. Zaĭtsev

St. Petersburg State Polytechnical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia
e-mail: guzenko@radio.rphf.spbstu.ru
Received February 17, 2003

Abstract—Multilayer wave-guiding structures comprising transversely magnetized ferrites are studied. A numerical theoretical model is constructed with the Galerkin method. An experimental investigation technique is developed. Theoretical and experimental results are found to be in good agreement. Such structures offer a high phase activity and may provide a basis for small-size millimeter-wave phase shifters and a new class of antenna systems, namely, integrated phased arrays. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

This paper is devoted to the theoretical and experimental investigation of a transversely magnetized three-layer ferrite–dielectric waveguide (FDW) whose cross-sectional view is shown in Fig. 1.

The waveguide consists of a high-permittivity dielectric slab sandwiched in ferrite slabs, of which the bottom of one is plated. Thus, the waveguide is an open guiding structure that can propagate waves with their fields concentrated mostly in the dielectric slab.

To control the magnetization of the ferrite layers, a control winding (not shown in Fig. 1) is provided between the ferrite slabs. The variation of the current in this winding alters the transverse magnetization of the ferrite layers and affects waves propagating in the FDW. Therein lies the essence of the electric control of the waveguide properties.

This waveguide has served as the basis for a new promising integrated millimeter-wave phased array with simple ferrite control elements [1, 2]. Here, the FDW functions as an integrated power and phase distributor. A control current changes the ferrite layer magnetization, which, in turn, alters the propagation constant of a wave in the FDW and, consequently, the phases of the radiators fed from it. This array thus offers simple electric steering of the beam by changing the current in the control winding.

A length of a layered FDW can be used as a small-size easy-to-control millimeter-wave phase shifter with losses lower than in conventional controlled semiconductor phase shifters.

The simulation of the FDW behavior is extremely important for the understanding of the physical mechanisms behind the processes taking place in an integrated phased array and in a controlled phase shifter.

The purpose of this work is (i) to create an adequate theoretical model of the FDW that could be used to study all types of waves existing in it, (ii) to estimate the controllability of the waves by changing the magnetization of the ferrite layers, and (iii) to optimize the characteristics of the structure, i.e., in terms of controllability maximization and loss minimization by appropriately choosing the number of layers and layer parameters (thickness, width, and permittivity). Such a model would eliminate the need for the experimental selection of these parameters, which requires the fabrication of many prototypes.

NUMERICAL SIMULATION

The theoretical simulation of an FDW is a complex electrodynamic problem, because it has a variable cross section and contains magnetized ferrites, which are anisotropic nonreciprocal media. For a structure with equiwide layers separated by electric or magnetic boundaries, an analytical solution can be constructed in principle (by applying the method of equivalent transmission lines). The resulting formulas are, however, very awkward and hard to use in practice. We study a waveguide with layers of different widths (Fig. 1), for
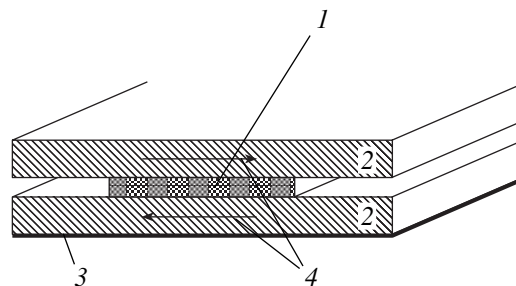

Fig. 1. Cross-sectional view of a transversely magnetized three-layer ferrite–dielectric waveguide: (*1*) dielectric, (*2*) ferrite, (*3*) metal, and (*4*) magnetization.

which an analytical solution cannot be constructed. Thus, we can rely only on a numerical method.

If a wave depends on time t and longitudinal coordinate z as $\exp[j(\omega t - \gamma z)]$, the Maxwell equations for the complex amplitudes of the fields take the form

$$\operatorname{curl}\mathbf{E} - j\gamma\mathbf{z}_0 \times \mathbf{E} = -j\omega\mu_0\overset{\leftrightarrow}{\mu}\mathbf{H},$$
$$\operatorname{curl}\mathbf{H} - j\gamma\mathbf{z}_0 \times \mathbf{H} = j\omega\varepsilon_0\varepsilon\mathbf{H}. \tag{1}$$

Here, $\overset{\leftrightarrow}{\mu}$ is the ferrite permeability tensor at a microwave frequency [3]. In the frequency range away from ferromagnetic resonance (such is the case in this paper, because the bias in an integrated phased array is weak), it can be represented as

$$\overset{\leftrightarrow}{\mu}\mathbf{H} = \mu\mathbf{H} - j\mu_a\mathbf{H} \times \mathbf{m_0}, \tag{2}$$

where $\mathbf{m_0}$ is the unit magnetization vector, which coincides here with $\mathbf{y_0}$, and μ and μ_a are, respectively, the diagonal and off-diagonal elements of the permeability tensor [3].

Thus, a desired solution is found by solving the set

$$\operatorname{curl}\mathbf{E} - j\gamma\mathbf{z}_0 \times \mathbf{E} + j\omega\mu_0(\mu\mathbf{H} - j\mu_a\mathbf{H} \times \mathbf{y_0}) = 0,$$
$$\operatorname{curl}\mathbf{H} - j\gamma\mathbf{z_0} \times \mathbf{H} - j\omega\varepsilon_0\varepsilon\mathbf{E} = 0. \tag{3}$$

The simulation of an FDW was performed with the Galerkin method [4, 5], which is a particular case of the method of moments. This method represents the components of a desired electromagnetic field as expansions in sets of linearly independent basis functions:

$$E_\nu(x, y) = \sum_{n=1}^{N}\sum_{m=1}^{M} a_{mn}^{(\nu)}\varphi_{mn}^{(\nu)}(x, y),$$
$$H_\nu(x, y) = \sum_{n=1}^{N}\sum_{m=1}^{M} b_{mn}^{(\nu)}\psi_{mn}^{(\nu)}(x, y); \quad \nu = x, y, z. \tag{4}$$

The basis functions $\varphi_{mn}^{(\nu)}(x, y)$ and $\psi_{mn}^{(\nu)}(x, y)$ must exactly satisfy the boundary conditions at the outer boundary (cross section) of the waveguide and must constitute complete sets, which is necessary to represent a desired solution.
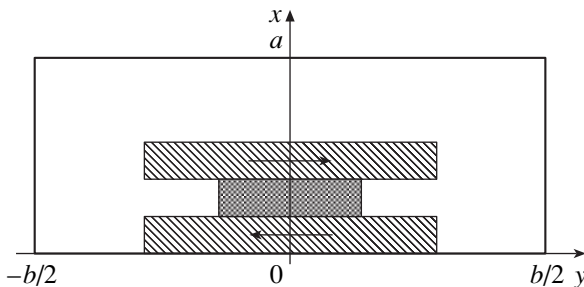

Fig. 2. Open FDW enclosed in an auxiliary screen.

Because an FDW is actually an open structure (Fig. 1), the boundary conditions at its boundary are conditions at infinity. Basically, sets of basis functions that satisfy these boundary conditions exist, but they are awkward and inconvenient for computations. Therefore, for simplicity, we construct the model for a completely screened FDW: an open FDW is enclosed in a large hollow rectangular waveguide (Fig. 2). This may produce extra modes, cavity modes, which have nothing to do with modes in a real open waveguide. However, the mode fields in an open FDW concentrate inside and near the layers and decay exponentially with distance from the waveguide; therefore, far screens are bound to have a negligible effect on the fields of sought modes, while the additional spurious (or "false") modes are expected to be very sensitive to the positions of the screens. This property is used to distinguish between true and false modes: when the positions of the auxiliary screens are varied, the cavity modes are significantly modified, while the modes of the open FDW retain their parameters.

The distances to the auxiliary screens should be optimized. On the one hand, the farther the screens, the weaker their effect on the modes of the open FDW and the more accurately these modes can be calculated; on the other hand, the greater number of basis functions are required in this case to represent the components of the desired modes (the field variation within the sheath becomes sharper). Accordingly, the calculation time and the number of false modes will increase considerably. Thus, optimum distances to the screens are those at which the parameters of true modes are calculated with a desired accuracy (they remain almost unchanged with a further increase in distance), and the detrimental factors discussed above moderate. This optimum can be found if the problem is solved by iterations: if the parameters of desired modes do not change (within a given accuracy) starting with a certain distance to the screens, this distance is fixed and then used in subsequent analysis of the structure.

The boundary conditions at the new closed boundary are very simple, and basis functions are easy to choose. In this paper, we expand the components of the desired field in trigonometric functions:

$$\left\{\begin{array}{c}\sin\\\cos\end{array}\right\}\frac{n\pi x}{a}\left\{\begin{array}{c}\sin\\\cos\end{array}\right\}\frac{m\pi y}{b}$$

An appropriate combination of the sine and cosine functions are determined for each of the field components from the corresponding boundary condition. These basis functions make up a complete set (as trigonometric functions) and are very convenient for calculations.

According to the Galerkin method [4, 5], expansions (4) are substituted into the left-hand side of field equations (3) written componentwise. The resulting expressions are multiplied by the respective basis func-

tions, integrated over the waveguide's cross section, and set equal to zero. As a result, we obtain a system of homogeneous linear algebraic equations for the expansion coefficients. The system contains the propagation constant as a free parameter. The nontriviality condition for a solution to this system yields the propagation constants (eigenvalues) of the FDW modes, for which the corresponding sets of expansion coefficients (eigenvectors) are calculated. With the expansion coefficients known, formulas (4) specify the field at any point of the waveguide and, hence, any integral characteristic of the mode can be calculated.

Based on this algorithm, a computer program that simulates an FDW and similar structures was developed. This program uses the waveguide parameters (layer dimensions, permittivities and loss tangents of the media, and ferrite magnetization) and the frequency of the oscillator as input parameters, constructs a relevant set of equations, solves it, and determines all necessary mode characteristics from the expansion coefficients found. The program is versatile: it applies to structures with any number of layers and with arbitrary widths and thicknesses of the layers. Also, it allows one to calculate the characteristics of many structures like an FDW (Fig. 1), which makes it particularly valuable.

EXPERIMENTAL TECHNIQUE

To experimentally study FDWs and similar structures, we used a special computerized measuring facility (CMF). The CMF is intended for measuring the amplitude–phase distribution along a straight line near the aperture (surface) of the object under study. The CMF is equipped with a miniature dipole probe, which introduces small perturbations into the field being measured. The software allows one to measure the amplitude–phase distribution and check the measurement accuracy (calibration mode), as well as display, print out, and file the data. The block diagram of the CMF is shown in Fig. 3. Probe *1* moves along object *3* by means of worm gearing *2*. Motor *4*, which actuates the mechanism, is controlled by computer-based unit *5*. The travel of the probe is specified by limit switches *6*. Marker generator *7* produces pulses that are uniquely related to the angle of rotation of drive shaft *2*, which makes it possible to locate the probe with an accuracy of ±0.1 mm. The amplitude and phase of the field are measured by amplitude–phase meter *8*, to which a signal from signal source *9* is applied. The signal from the movable probe is transmitted to the amplitude–phase meter through cable *10*, a Teflon-filled coaxial transmission line of outer diameter 1.5 mm. This cable is flexible and lets the probe move over noticeable distances. The electrical length of the cable remains constant within several degrees. Interface device *11* controls the probe motion, the operation of the amplitude–phase meter, and the entry of the data into a computer.

The basic parameters of the CMF are as follows: the frequency range is 32.0–35.7 GHz (beyond this range,
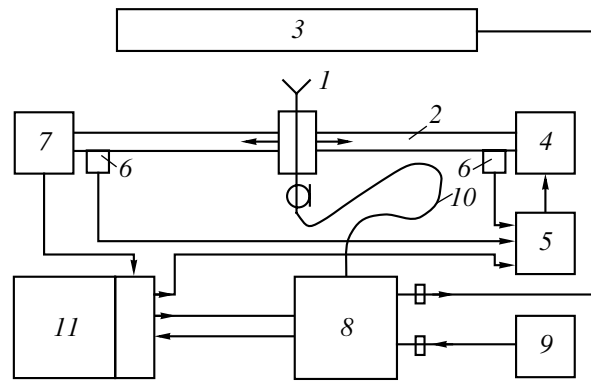

Fig. 3. Block diagram of the measurement facility.

the measurement accuracy is lower), the amplitude–phase distribution is measured to within ±3° (phase) and ±0.5 dB (amplitude), the length of measurement is varied from 0.8 to 1200 mm, the minimum distance between amplitude and phase counts is 0.1 mm, and the probe positioning error is within ±0.1 mm.

A length of the FDW was placed into the CMF. The probe moved ≈0.1 mm above the FDW surface so that it could respond to waves propagating in the FDW. The waveguide was driven through a narrow slot, which efficiently excites a variety of modes.

The output data of the CMF are the readings of the field amplitudes and phases at given points of the FDW surface. To calculate the FDW mode characteristics of interest, a special data processing technique and associated software package were developed.

The amplitude and phase readings form readings of the field complex amplitude. Then they are subjected to Fourier transformation (in the space domain). The theory predicts that the propagation constants of FDW modes are discrete and that any wave in the waveguide is the superposition of these modes. Thus, the Fourier spectrum has isolated peaks, which correspond to modes that exist in a particular case (Fig. 4 shows such a spectrum for a frequency of 36 GHz). The parameters of these modes (propagation constant and attenuation factor) can be found from the spectrum with a fairly good accuracy.

For each mode of the spectrum is transformed so that only a part of the spectrum near the peak corresponding to this mode is left, and the remaining part of the spectrum is set equal to zero. Next, the new (truncated) spectrum is subjected to inverse Fourier transformation; i.e., the amplitude–phase distribution of the mode is obtained. It is split into the amplitude and phase distributions. From the amplitude distribution (on the logarithmic scale), the attenuation factor can easily be calculated by using the linear approximation. The phase distribution is used to calculate the propagation constant (also with the linear approximation). This technique gives the propagation constant accurate to 0.1–0.3%. Thus, taking simple measurements, one may
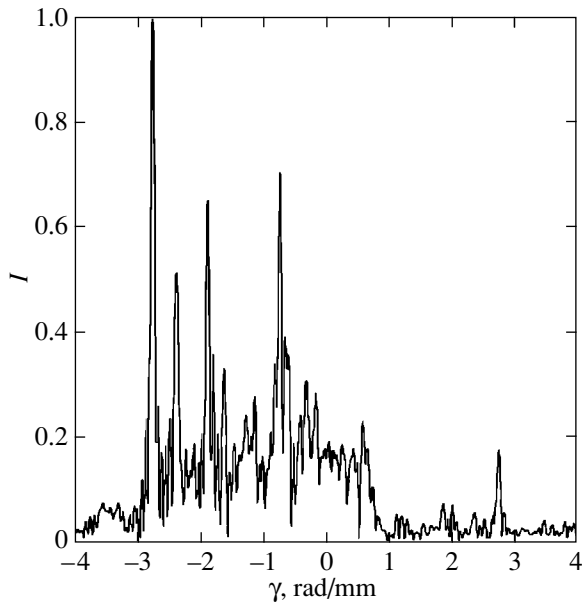
Fig. 4. Fourier spectrum of the amplitude–phase distribution.

for measuring the coupling coefficient between FDW modes and radiators. The analysis of the amplitude–phase distribution in different regions of the FDW surface allows us to estimate the degree of FDW longitudinal inhomogeneity, which is caused by fabrication errors. The modes reflected from the terminating load also appear in the spectrum and have a propagation constant of opposite sign. Analyzing the reflected waves, one can select a load that reflects a particular mode. The above technique can also be used in other investigations.

This technique was implemented as a software package for processing experimental data and conducting various studies.

RESULTS AND DISCUSSION

The theoretical and experimental methods described above apply to a variety of structures similar to a ferrite–dielectric waveguide. Below, we discuss the results for the three-layer waveguide whose cross-sectional view is given in Fig. 1, because integrated phased arrays are built on this waveguide.

Figure 5 shows the slowing factor (the propagation constant normalized to the wave number in free space) for modes of the demagnetized FDW versus frequency over a sufficiently wide frequency range. Shown are the theoretical curves and data points. The calculated and measured results are seen to be in excellent agreement for the fundamental mode (the one with the highest slowing factor). For modes with slowing factors of 2.0–3.5, agreement is also good. Even two modes with close values of the slowing factor (open circles) are resolved both theoretically and experimentally. For the highest

gain very important information about modes that exist in the waveguide under study.

The technique described above may provide data for other important properties of the waveguide. In a real integrated phased array, it is preferable to excite only the fundamental mode (which has the lowest phase velocity) at the FDW input. A theoretical model for an FDW driver is still lacking. However, the distribution of mode peak heights in the spectrum may be used to optimize the driver. We also intend to extend this technique
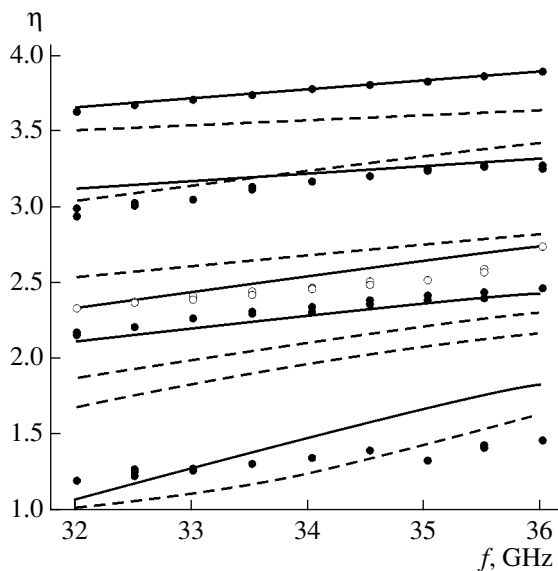

Fig. 5. Slowing factor versus frequency for symmetric (solid lines) and antisymmetric (dashed lines) FDW modes. Symbols, data points.
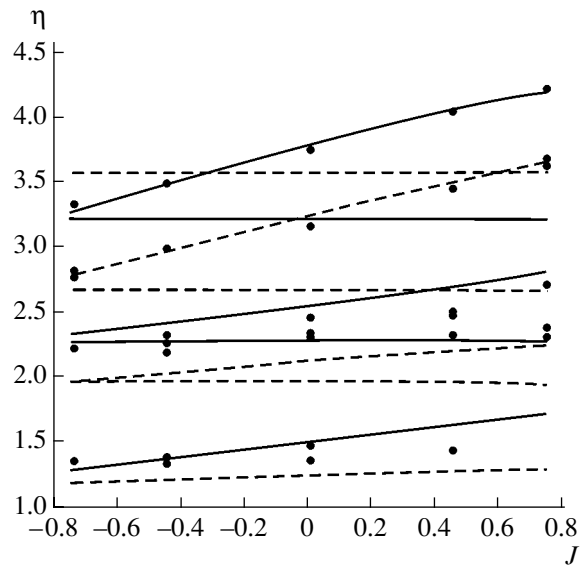

Fig. 6. Calculated controllability of symmetric (solid lines) and antisymmetric (dashed lines) FDW modes. Symbols, data points. J is the relative magnetization, and η is the slowing factor.

order mode (the one with the lowest slowing factor), agreement is worse; however, the very fact that the presence of this mode is predicted by the theory counts in favor of the model constructed. A discrepancy between the calculated and experimental results is inevitable, because it is impossible to allow for all specific features of a real FDW, such as the presence of the control winding and magnetic-flux-closing elements at the ends of the structure, etc. [1, 2]. Modes with less-than-unity slowing factors cannot exist in an open FDW. In our case, they are leaky waves or radiation modes of the open structure. These modes are very sensitive to the screen position (Fig. 2).

Figure 6 shows the slowing factor for various FDW modes versus relative magnetization (normalized to the saturation magnetization) of the ferrite layers at 34 GHz. The magnetization interval in Fig. 6 is feasible when the current in the control winding varies within ±1 A. As can be seen in Fig. 6, different FDW modes show different degrees of the controllability by the ferrite magnetization. Some of them are uncontrollable. Of most interest is the fundamental FDW mode, which has the highest amplitude and is the operating mode in integrated phased arrays [1, 2]. Its slowing factor varies from 3.27 to 4.18 (24%). For an integrated phased array, this means that the beam scans a sector of about 40° [1, 2]. If a segment of the FDW is used as a phase shifter, its length must be about 1.1 times as large as the wavelength to provide a phase shift of 2π.

Such a high phase activity of this medium is due to its layered structure, which consists of the two oppositely magnetized ferrite layers and the dielectric layer. The dielectric transforms the FDW field so that the magnetic field components H_x and H_z are close in amplitude and phase-shifted by $\pi/2$ (i.e., rotate about the constant magnetization direction) throughout the ferrite-filled space.

It should be noted that the modes found theoretically were not all observed in the experiment (like the peaks in the spectrum), because they are weakly excited at the waveguide input and/or weakly coupled with the probe.

Since the waveguide has a symmetric cross section (Fig. 1), its modes can be subdivided into symmetric (the field tangential component is maximum at the waveguide axis) and antisymmetric (the tangential component at the axis is zero) modes. The FDW was excited by a standard rectangular waveguide with a symmetric mode and had a symmetric waveguide transformer at the input. Therefore, we first assumed that only symmetric modes are excited. However, Fig. 6 shows that at least one antisymmetric mode (the second controllable mode) is excited in the experiment. This can be explained by the asymmetry of the waveguide transformer of the FDW prototype.

CONCLUSIONS

In this paper, we theoretically and experimentally studied a three-layer ferrite–dielectric transversely magnetized waveguide. A numerical model of the FDW is constructed based on the Galerkin method. A dedicated measuring and data-processing technique is developed. The calculated and experimental data are in good agreement, which testifies that the model is adequate to a real physical system.

It is shown that different waveguide modes can be controlled to a variable degree. When the ferrite layer magnetization varies over the feasible range, the propagation constant of the fundamental operating mode, which is of most interest, varies by 24%. This means that the phase activity of the medium is sufficiently high due to its inhomogeneity (layered structure).

The model constructed can be used to study waves propagating in a variety of ferrite–dielectric structures and to optimize the characteristics of the structure, for example, to maximize the controllability by appropriately choosing the number and parameters of the layers. This eliminates the need for fabricating a great number of prototypes in experiments.

Our results demonstrate that transversely magnetized ferrite–dielectric structures are promising for electrically controlled millimeter-wave integrated scanning antennas and small-size phase shifters.

REFERENCES

1. E. F. Zaitsev, Y. P. Yavon, Y. A. Komarov, *et al.*, IEEE Trans. Antennas Propag. **42**, 304 (1994).

2. Y. S. Khodorkovsky, E. F. Zaitsev, A. B. Gouskov, *et al.*, in *Proceedings of the Symposium on Antenna Application, Monticello, 1997.*

3. A. L. Mikaélyan, *Theory and Application of Ferrites at Microwaves* (Gosénergoizdat, Moscow, 1963).

4. V. V. Nikol'skiǐ, *Variational Methods for Inner Problems of Electrodynamics* (Nauka, Moscow, 1967).

5. G. I. Marchuk, *Methods of Numerical Mathematics* (Nauka, Moscow, 1980; Springer, New York, 1975).

Translated by A.D. Khzmalyan

ELECTRON AND ION BEAMS, ACCELERATORS

Measurements of the Transverse Electron Velocities in High-Current Microsecond Relativistic Electron Beams in a Strong Magnetic Field

O. T. Loza and I. E. Ivanov

Prokhorov Institute of General Physics, Russian Academy of Sciences,
ul. Vavilova 38, Moscow, 119991 Russia
e-mail: www.gpi.ru
Received June 4, 2002; in final form, November 5, 2002

Abstract—An analyzer is created for time-resolved measurements of the electron pitch-angles in high-current microsecond relativistic electron beams in a strong magnetic field. The electron pitch-angles in a 500-keV relativistic electron beam with a current density of ~1 kA/cm^2 and a 1-μs flat-top current profile are measured. The diode proposed previously by the authors allows one to produce a high-current electron beam in which pitch-angles vary only slightly with time and over the beam cross section. © 2003 MAIK "Nauka/Interperiodica".

INTRODUCTION

High-current relativistic electron beams (REBs) is an accepted definition for electron beams with energies of ~10^6 eV and electron current densities of ~10^3–10^4 A/cm^2. The REB power usually exceeds 10^9 W, and the current pulse lasts from several nanoseconds to several microseconds. The high power of REBs makes them attractive for various applications; however, a rather short REB duration causes difficulties for diagnosing the beam parameters.

High-current REBs are usually formed and propagate in a magnetic field with an induction of ~1 T or higher. In such a field, an electron moves along a helical trajectory and its velocity is directed at an angle Θ to the magnetic field. The angle Θ between the longitudinal (i.e., directed along the magnetic field) component of the electron velocity v_l and its transverse component v_t is called the pitch-angle ($\tan \Theta = v_t/v_l$).

High-current REBs are usually generated with the help of explosive emission cathodes. For current pulse durations of about a few microseconds, the plasma boundary from which electrons are emitted shifts away from the cathode surface by several centimeters [1], so that the parameters of the electron trajectories in a strong magnetic field at the beginning and the end of the pulse can be substantially different. The problem of measuring the electron pitch-angles in microsecond REBs formed with an explosive emission cathode has up to now been inadequately studied, and difficulties encountered in such measurements increase with decreasing pitch-angle.

There are a number of methods for measuring pitch-angles in high-current REBs. A diamagnetic probe [2] provides data only on the mean value of the pitch-angle of electrons at a given instant of time. The character of

the electron angular distribution function can be judged from measurements of the trajectories of a small fraction of electrons that are, in some way, separated from the beam. These (contact) methods usually employ either relatively long large-diameter cylindrical channels [3], which are also known as "wells," or relatively short (along the axis) small-diameter apertures (pinholes) [4].

A common disadvantage of all the contact methods used in the physics of high-current microsecond REBs is the production of a parasitic plasma on the detector components. The threshold density for the plasma generation is ~1 J/cm^2; hence, the plasma arises on any collector several tens of nanoseconds after the beginning of the pulse. This plasma has a density of 10^{12}–10^{13} cm^{-3} and propagates along the magnetic field lines toward the electron beam with a velocity higher than 10^7 cm/s [5]. The effect of such a dense plasma on the angular characteristics of the REB electrons can amount to several tens of degrees [4]. Thus, an analyzer with an aperture cutting a fraction of the beam for the subsequent diagnostics can substantially affect the beam parameters to be measured.

At lower REB current densities, the plasma is generated at a slower rate. Since the external magnetic field B substantially exceeds the self-magnetic field of the electron beam, a decrease in the current density can be achieved by decreasing the guiding magnetic field. Unfortunately, the self-electrostatic field of a high-current REB limits the applicability range of this method.

For an electron beam with a circular cross section and radius R_b, we have $BR_b^2 = $ const along its pathway. The self-electrostatic field on the beam surface is inversely proportional to the beam radius, $E \sim R_b^{-1}$; i.e.,

$E \sim \sqrt{B}$. In crossed electric field **E** and magnetic field **B**, an electron drifts with the velocity $v_{\mathrm{dr}} = cE/B$; i.e., $v_{\mathrm{dr}} \sim B^{-1/2}$. The drift velocity v_{dr} of an electron is added to its oscillatory velocity v_t; as a result, the pitch-angle $\Theta_0 = \arctan(v_t/v_l)$ varies periodically (with the electron gyrofrequency) within the range determined by the relationship $\tan\Theta \approx (v_t \pm v_{\mathrm{dr}})/v_l \equiv \tan(\Theta_0 + \Delta\Theta)$. The weaker the magnetic field near the analyzer, the greater the relative error in measuring the pitch-angle, $\Delta\Theta/\Theta_0 \sim v_{\mathrm{dr}}/v_t \sim B^{-1/2}/B^{1/2} \sim 1/B$. As a result, at sufficiently low magnetic fields, the measurements can become meaningless. For example, in [6], a pitch-angle of $\Theta \approx 40°$ was measured in a gyrotron with the REB radius $R_b = 9$ mm, electron energy 350 keV, current $I = 6$ kA, and magnetic field 2 T. It can easily be estimated that, in this case, the pitch-angle periodically varies within $\approx 20\%$. As the magnetic field is decreased fivefold, the scatter in the electron pitch-angles, $\Delta\Theta/\Theta_0 \sim 1/B$, becomes so large that the electron beam turns out to be unusable for measurements.

This paper is aimed at solving two problems. The first problem is to develop a method for measuring the distribution of the beam electrons over pitch-angles (down to pitch-angles as small as possible). The method should have high angular resolution in a strong magnetic field and should be insensitive to the influence of the parasitic plasma. The second problem is to verify whether the angular distribution of electrons in a microsecond REB generated with an explosive emission cathode is the same at the beginning and the end of the pulse.

OPERATION OF THE PITCH-ANGLE ANALYZER

The analyzer proposed in this paper is based on the pinhole method. Figure 1 shows the operating principle of such an analyzer. An electron (*1*) with a velocity v enters a measuring chamber through an aperture in a diaphragm (*2*) and falls on a screen (*3*) at a certain distance r from the aperture axis. The distance between the diaphragm and the screen is L, and the angle between the electron velocity and the normal to the diaphragm is Θ. The electron pitch-angle is deduced from the distance r between the axis and the position of the electron on the screen. The recording of electrons may be either integral over the entire pulse (the photographing of the glow of a mylar film [4] or a ZnS plate [7]) or time-resolved (a small-size sectioned collector consisting of several annular electrodes [8]).

In order for the distance r from the electron position on the screen to the axis to be uniquely determined by its pitch-angle Θ, the distance L between the diaphragm and the screen should be substantially smaller than the spatial period λ_H of the helical electron trajectory in the magnetic field: $L \ll \lambda_H$. In this case, the electron trajec-
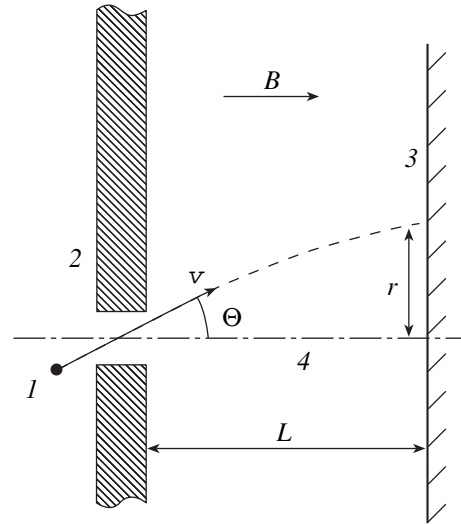

Fig. 1. Analyzer of the angular spectrum: (*1*) electron, (*2*) diaphragm, (*3*) screen, and (*4*) symmetry axis.

tory in the analyzer is close to linear, so that $r \cong L\tan\Theta$. For moderate Θ values, we have

$$L \sim \frac{\lambda_H}{2\pi} = \frac{v_l}{\Omega_H}\gamma \approx \frac{c\sqrt{\gamma^2 - 1}}{\Omega_H}$$
$$\approx \frac{1.7 \times 10^{-3}[\text{T m}]}{B[\text{T}]}\sqrt{\gamma^2 - 1}. \qquad (1)$$

Here, Ω_H is the nonrelativistic electron gyrofrequency, γ is the relativistic factor, c is the speed of light, and B is the induction of the external magnetic field. For $\gamma = 2$ and $B = 1$ T, condition (1) takes the form $L < 3$ mm and the electron with a pitch-angle of $\Theta = 10°$ falls at the distance $r \cong 0.5$ mm from the axis, so that its detection involves difficulties. Hence, the presence of a strong magnetic field in the pin-hole analyzer is a factor limiting the device length and adversely affecting the accuracy of measurements.

This limitation can be avoided if the magnetic field rapidly (at a distance Δz) drops to zero near the analyzer diaphragm. If Δz is sufficiently small, an electron retains information about its pitch-angle in the magnetic field and continues moving inside the analyzer along a linear trajectory. In this case, the analyzer length may be long enough to achieve the desired resolution.

The desired distribution of the magnetic field can be achieved by applying an additional field compensating for the spatially nonuniform field B_0 in the analyzer. Figure 2 shows a conical surface with a radius R, which varies along the z axis as $R = z\tan\alpha$. Let circular currents with a certain distribution flow along the conical surface and produce inside the cone a magnetic field that has the same magnitude as the external field B_0 but is opposite in sign. Let us estimate the magnetic field to
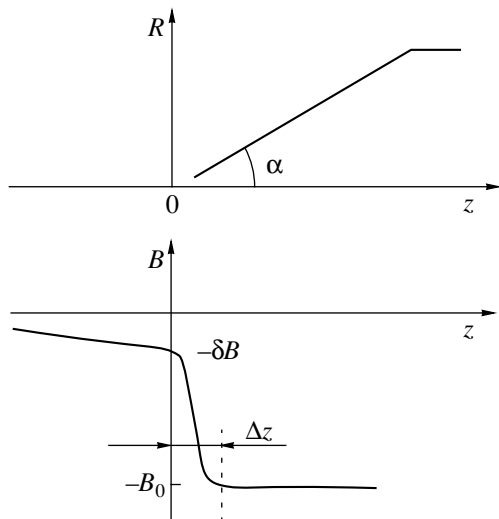
Fig. 2. Profile of the analyzer surface (top) and the longitudinal profile of the compensating magnetic field (bottom).
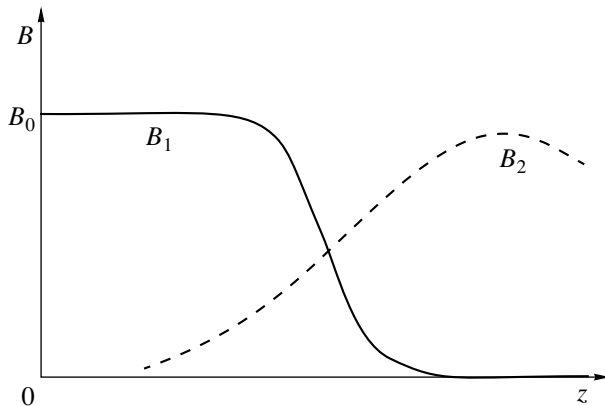

Fig. 3. Magnetic field outside the analyzer: B_1 is the basic quasi-steady field and B_2 is the pulsed field.

the left of the cone, assuming for simplicity that the aperture in the diaphragm is absent. The magnetic field produced by a thin-wire circular loop with a radius R, current I, and coordinate z on the axis at the point with the coordinate z_1 is equal to

$$B(z, z_1) = \frac{IR^2}{\left(R^2 + (z - z_1)^2\right)^{3/2}}$$

(a factor depending on the specific system of units is omitted).

At $z = z_1$, i.e., in the center of the loop, the field is $B = I/R$, and at the point $z_1 = 0$, it is

$$B(z, 0) = \frac{IR^2}{\left(R^2 + z^2\right)^{3/2}}$$

$$= \frac{IR^2}{\left(R^2 + (R\cot\alpha)^2\right)^{3/2}} = \frac{I}{R}\frac{1}{\left(1 + \cot^2\alpha\right)^{3/2}}.$$

For $\alpha = 20°$, we have $(1 + \cot^2\alpha)^{-3/2} \approx 0.04$. Since this consideration is true for any loop with any z coordinate, we can conclude that circular currents flowing along the conical surface distort the magnetic field to the left of the cone (at $z \leq 0$) by no larger than 4%: $\delta B \approx 0.04B_0$ (Fig. 2). The length Δz of the transition region is comparable in size with the diameter of the aperture situated at the point $z \approx 0$.

The desired profile of the magnetic field can be produced using the skin effect in a pulsed field. The magnetic field is a superposition of two fields: the basic field B_1 and the field in the analyzer B_2 (Fig. 3). A necessary requirement is that the basic field (it may be stationary) be zero ($B_1 = 0$) inside the analyzer. The field B_2 is pulsed, so it does not penetrate into the analyzer, but outside it is rather strong. The total field $B_1 + B_2$ should be nearly constant along the axis: if the magnetic field increases substantially along the electron trajectory, then the electrons are reflected from the magnetic mirror, whereas if the field decreases, then the self-electric field of the beam comes into play. Such a nearly uniform magnetic field ensures the adiabatic motion of electrons (which is necessary to retain information on their pitch-angle) to the analyzer.

The configuration of the magnetic field lines is determined by the skin effect, due to which the magnetic field lines are expelled from the analyzer. As a result, inside the analyzer, the field is absent, whereas outside it, where the magnetic field induction remains large, the field lines diverge along the conical surface of the analyzer. Hence, the magnetic field makes the electron trajectories diverge, so that only electrons moving near the axis can pass through the aperture. Most of the electrons moving along the field lines avoid collisions with the analyzer surface, thus preventing the production of a dense plasma. Of course, some electrons bombard the aperture edges; however, their number is much less than in a uniform field. Let us emphasize once again that our consideration is valid if the spatial period of the electron helical trajectory λ_H is much larger than the length Δz of the transition region.

Therefore, the analyzer operates as follows. The REB electrons pass through the aperture into the analyzer at an angle equal to the pitch-angle of their trajectories in the magnetic field. In the analyzer, where the magnetic field is absent, they propagate along straight lines until they meet a detector (e.g., scintillator). The higher the detector sensitivity, the larger the maximum distance between the detector and the diaphragm and the higher the resolution in pitch-angles. Note that it is the aperture diameter that determines the number of electrons penetrating through the diaphragm and, accordingly, the maximum distance to the detector.

DESIGN AND CALIBRATION OF THE DEVICE

The analyzer is designed as follows. The 3-mm-thick analyzer casing is a cone that transforms into a cylinder ≈50 mm in diameter. The casing is made of aluminum and copper. The angle between the cone generatrix and the axis is $\alpha = 20°$, and the distance between the 1-mm-diameter aperture and the detector is 60 mm. In time-integrated measurements, an astralon film, which changes its color under the action of electrons, was used as a detector. In order to determine how the properties of the beam electrons change with time, we used a radially sectioned current collector similar to that used in [8]. Seven coaxial annular collectors distributed along the radius allowed us to resolve electron pitch-angles with a step of ≈2.3°.

The analyzer was calibrated by a 500-keV electron beam with a current density of up to 1 kA/cm^2 and pulse duration of 1 μs. The electrons propagated in a quasi-steady uniform magnetic field with an induction of $B_1 = 1.1$ T. Over a distance of ~20 cm, the field B_1 fell to almost zero (<0.04 T) in the region where the analyzer was installed. Over the same distance of 20 cm, the pulsed field increased to 0.8 T. For a pulsed field duration of 30 μs, the skin-layer thickness in copper is 0.5 mm.

We found conditions under which most of the measured electron trajectories had pitch-angles less than 2° throughout the entire REB current pulse (as was deduced from the degree of blackening of the astralon film). Under these conditions, the device was calibrated by the scattering of the beam electrons by a 20-μm aluminum foil. The foil was placed nearly in the middle of the gap between the diode and the analyzer, at a distance of ~0.5 m from each of them.

It is well known that, when electrons with the same velocities are elastically scattered by a foil, the distribution of the electron current density J over the scattering angles ϑ is Gaussian, $J(\vartheta) \sim \exp(-\vartheta^2/\vartheta_0^2)$. Here, ϑ_0 is the mean square deviation, which depends on the foil characteristics and the electron energy [9]. For the given foil and electron energy, we have $\vartheta_0 \approx 10°$.

For the given dependence of the current density $J(\vartheta)$ (which determines the blackening profile of the astralon film with a maximum at $\vartheta = 0$), the distribution of the total electron current I over the angle ϑ (i.e., the current density profile measured by the sectioned collector) has the form

$$I(\vartheta) \sim \sin\vartheta \exp(-\vartheta^2/\vartheta_0^2).$$

Figure 4 shows the electron current measured by each of seven collectors (i.e., the current density within seven angular ranges). These currents are compared with the calculated data, i.e., with the signals expected for the scattering of monoenergetic electrons with velocities strictly parallel to the magnetic field. The calculations were performed taking into account the
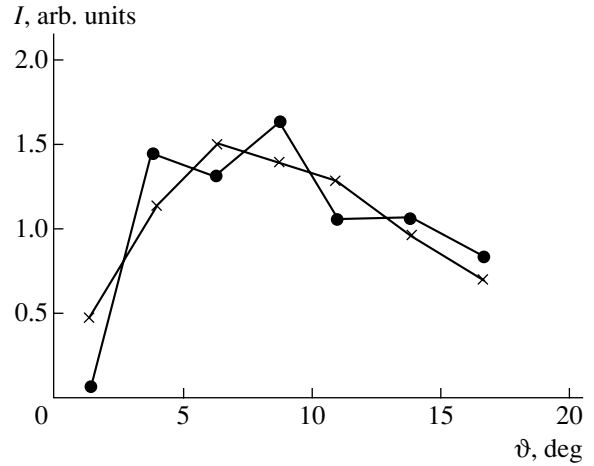

Fig. 4. Calculated (×) and measured (●) current I produced by the electrons with different pitch-angles ϑ.

instantaneous electron energy, measured by the cathode potential detector.

The collector signals were treated using the least squares method. The experimental values of the electron current $I(\vartheta)$ were approximated by the function

$$\varphi(\vartheta) = A \sin\vartheta \exp(-\vartheta^2/\vartheta_0^2),$$

where A and ϑ_0 are the normalizing factor and the root-mean-square scattering angle, respectively.

It can be shown that, for every instant of time, the normalizing factor is equal to

$$A = \left(\sum_{k=1}^{7} \varphi_k I_k\right)\left(\sum_{k=1}^{7} \varphi_k^2\right)^{-1},$$

and ϑ_0 satisfies the equation

$$\sum_{k=1}^{7} \varphi_k I_k \sum_{k=1}^{7} \vartheta_k \varphi_k^2 - \sum_{k=1}^{7} \varphi_k I_k \vartheta_k \sum_{k=1}^{7} \varphi_k^2 = 0.$$

Here, the summation is over all the collectors ($k = 1,…, 7$), ϑ_k is the mean angle for the scattering into the kth collector, I_k is the signal from the kth collector, and $\varphi_k = \varphi(\vartheta_k) = A \sin\vartheta_k \exp(-\vartheta^2/\vartheta_0^2)$.

On processing the data shown in Fig. 4, we can compare the r.m.s. deviation angles ϑ_0. The calculations by the formulas from [9] give $\vartheta_0 = 10.6°$, whereas the measured value is $\vartheta_0 = 11.6°$ with a variance of 0.6°. Some excess of the measured ϑ_0 value over the calculated one may be attributed to the small angular spread of the electrons before their interaction with the foil.

Figure 5 shows the time dependences of the calculated (by the formula from [9]) and measured r.m.s. scattering angle ϑ_0 of electrons after their interaction with the foil. Since the cathode potential, as well as the REB total current, changed insignificantly over 1.5 μs,
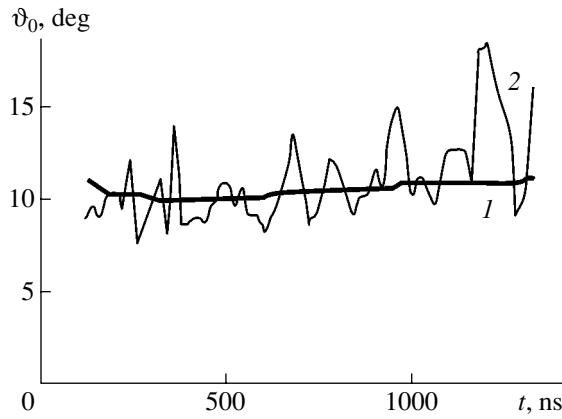
Fig. 5. Scattering angle ϑ_0 as a function of time t: (*1*) calculations and (*2*) the results of processing the analyzer data.
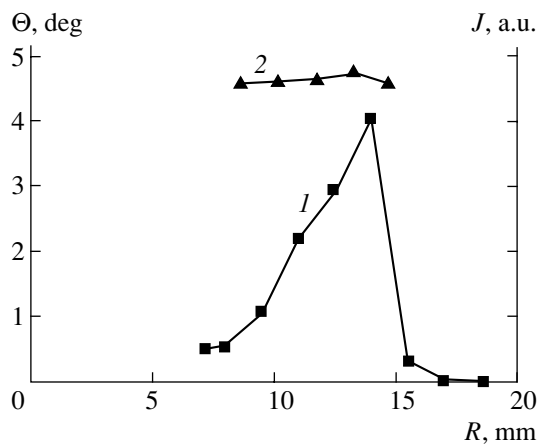

Fig. 6. Time-integrated distributions of the (*1*) electron current density and (*2*) maximum pitch-angles over the radius R.
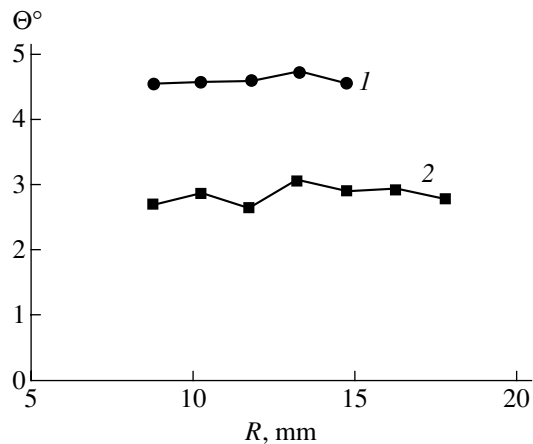

Fig. 7. Pitch-angle vs. R for two values of the magnetic field: (*1*) 1.1 and (*2*) 0.55 T.

the ordinates of the calculated points are nearly the same. The signals from the collectors as functions of time were treated by the procedure described above. The results obtained agree well with the calculated data.

The intensity of plasma production on the analyzer surface was experimentally estimated as follows. The conical surface of the analyzer was covered with a thin film of a colorant sensitive to electron bombardment and heating. After the multiple REB action on the device, the film became discolored only near the cone vertex. The maximum radius of the discolored part of the cone did not exceed 3 mm; i.e., it was on the order of the length calculated by formula (1). Hence, we believe that we have avoided an intense plasma production.

Thus, when measuring relatively large (~10°) pitch-angles, the sectioned collector has demonstrated the high measurement accuracy. However, the use of such a collector imposes rigid requirements on the device adjustment; therefore, it seems more convenient to use a scintillator combined with a high-speed image intensifier.

EXPERIMENT

Experiments on studying the electron trajectories were carried out with a high-current electron accelerator generating 500-keV electron beams with a pulse duration of ~1 μs. An annular electron beam with a current of 2 kA was formed in a diode [10] that ensured unchanged beam geometry on a microsecond time scale.

An annular electron beam 28 mm in diameter propagated in a quasi-steady uniform magnetic field along a 90-mm-diameter tube (with the limiting vacuum current exceeding 3 kA) over a distance of ~1 m. The analyzer was placed in the segment where the drop in the quasi-steady magnetic field was compensated for by a rapidly varying magnetic field. The magnetic induction in the region where the magnetic field was uniform was ≈1.1 T.

The REB current density as a function of radius is shown in Fig. 6. The current distribution shows that the electron beam is hollow; its outer radius is ~15 mm, and its thickness is 3 mm. The figure also shows the time-integrated radial distribution of the electron pitch-angles ϑ, which was recorded with the help of an astralon film. One can see that, throughout the cross section of the electron beam, in which the current density varies by one order of magnitude, the electron pitch-angles are approximately the same and do not exceed 5°.

Some experiments were performed with a decreased magnetic field. In this case, the pulsed magnetic field in the analyzer was also decreased, so that the shapes of the magnetic field lines did not change. The radial distributions of the electron pitch-angles ϑ in an REB in a magnetic field with an induction of 1.1 and 0.55 T are shown in Fig. 7. In the lower field, the electron beam has a somewhat larger outer radius (about 17 mm) and a larger thickness. In both the higher and lower field, the pitch-angles are almost the same throughout the

REB cross section; however, in the case of the lower field, the pitch-angles are somewhat lower and do not exceed 3°.

The time-integrated measurements, whose results are presented in Figs. 6 and 7, were carried out with a pulse duration of 1 μs. The pulse duration could be varied from 0.3 to 1.0 μs, and the experiments were also carried out for several pulse durations within this range. The results obtained differ only slightly from those shown in Figs. 6 and 7; for this reason, we can assert that the electron pitch-angles vary slightly throughout the entire microsecond current pulse.

CONCLUSIONS

An analyzer of electron pitch-angles in a strong magnetic field has been designed. The analyzer is based on the pinhole method. A small fraction of the beam electrons penetrates through a small aperture into a chamber, where their trajectories are analyzed. At the entrance to the device, the induction of the magnetic field sharply decreases to zero and the electrons continue moving freely in the analyzer with a constant velocity. One may use various recording techniques: scintillators, collectors, etc. The angular resolution depends on the distance between the aperture and the detector plane.

The required profile of the magnetic field is produced by superposing two fields. One of these fields is a basic quasi-steady magnetic field that vanishes inside the analyzer. The other field is pulsed and is added to the basic field for the total field to be spatially uniform. The pulsed field cannot, however, penetrate into the analyzer because of the skin effect. The advantages of the conical shape of the device casing is threefold: (i) the magnetic field profile in front of the analyzer is only slightly distorted, (ii) the magnetic field drops sharply at the entrance to the device, and (iii) most of the electrons move apart from the analyzer, thus preventing an intense plasma production.

The device calibration by the scattering of relativistic electrons by an aluminum foil have demonstrated that the experimental data agree well with the calculated results. The calibration has also shown that it is possible to measure the angular distribution of electrons in an REB with a current density of ~1 kA/cm^2 and to trace its time evolution at relatively small pitch angles of ~10°.

The electron pitch-angles in a 500-keV REB with a current of 2 kA have been measured. It is shown that with a diode designed previously in [10], it is possible to form a high-current annular electron beam in which the current-density profile remains uncharged on a microsecond time scale and in which the electron pitch-angles vary only slightly with time and over the beam cross section. As the magnetic field increases, the pitch-angles increase and amount to 5° in the magnetic field with an induction of 1.1 T.

REFERENCES

1. S. P. Bugaev, A. A. Kim, A. I. Klimov, and V. I. Koshelev, Fiz. Plazmy **7**, 529 (1981) [Sov. J. Plasma Phys. **7**, 286 (1981)].

2. Yu. I. Abrashitov, V. S. Koĭdan, V. V. Konyukhov, *et al.*, Zh. Éksp. Teor. Fiz. **66**, 1324 (1974) [Sov. Phys. JETP **39**, 647 (1974)].

3. A. V. Arzhannikov, V. S. Koĭdan, and S. V. Loginov, Prib. Tekh. Éksp., No. 4, 36 (1983).

4. P. S. Strelkov, A. G. Shkvarunets, and P. Shunka, Fiz. Plazmy **7**, 564 (1981) [Sov. J. Plasma Phys. **7**, 305 (1981)].

5. N. I. Zaĭtsev and G. S. Korablev, Zh. Tekh. Fiz. **52**, 160 (1982) [Sov. Phys. Tech. Phys. **27**, 110 (1982)].

6. V. I. Krementsov, P. S. Strelkov, and A. G. Shkvarunets, Zh. Tekh. Fiz. **50**, 2469 (1980) [Sov. Phys. Tech. Phys. **25**, 1447 (1980)].

7. G. Bekefi, R. E. Shefer, and S. C. Tasker, Nucl. Instrum. Methods Phys. Res. A **250**, 91 (1986).

8. P-H. De Haan, R. N. Singh, H. J. Hopman, *et al.*, J. Phys. E **14**, 373 (1981).

9. E. Segre, in *Experimental Nuclear Physics*, Ed. by E. Segre (Wiley, New York, 1953; Inostrannaya Literatura, Moscow, 1955), Vol. 1.

10. O. T. Loza and P. S. Strelkov, in *Proceedings of the 12th International Conference on High-Power Particle Beams, Haifa, 1998*.

Translated by E. Satunina

Improvement of the Efficiency of a Glow Discharge-Based Ion Emitter with Oscillating Electrons

N. V. Gavrilov and D. R. Emlin

*Institute of Electrophysics, Ural Division, Russian Academy of Sciences,
ul. Amundsena 106, Yekaterinburg, 620016 Russia*
e-mail: pulsar@iep.uran.ru
Received July 25, 2002

Abstract—Ion emission from the plasma of a low-pressure ($\approx 5 \times 10^{-2}$ Pa) glow discharge with electrons oscillating in a weak (≈ 1 mT) magnetic field is studied in relation to the cold hollow cathode geometry. A hollow conic cathode used in the electrode system of a cylindrical inverted magnetron not only improves the extraction of plasma ions to $\approx 20\%$ of the discharge current but also provides the near-uniform spatial distribution of the ion emission current density. The reason is the specific oscillations of electrons accelerated in the cathode sheath. They drift in the azimuth direction along a closed orbit and simultaneously move along the magnetic field toward the emitting surface of the plasma. A plasma emitter with a current density of ≈ 1 mA/cm² over an area of ≈ 100 cm² designed for an ion source with an operating voltage of several tens of kilovolts is described. © 2003 MAIK "Nauka/Interperiodica".

INTRODUCTION

The energy cost of an ion, $w = (I_d U_d)/I_i$, in ion beams generated by hollow-cathode low-pressure glow discharge sources is much higher than in those produced by thermionic-cathode sources and is usually several kilo-electron-volts per ion. The reason is the poor ion–electron emissivity of a cold cathode and, accordingly, the high discharge operating voltage U_d. Another problem is associated with the extraction of a considerable fraction of the discharge current ionic component ($\alpha \approx I_i/I_d$) from the plasma. The oscillations of emitted electrons, which are accelerated in the cathode sheath, provide a fairly uniform distribution of the discharge current over the hollow cathode surface. Therefore, the ratio of the ion current extracted from the plasma to the discharge current is roughly equal to the ratio of the emitting surface area of the plasma, S_e, to the total surface area of the cathode, S_c. In discharge systems with hollow cathodes of a nearly conic shape, the ratio $\alpha = S_e/S_c$ as high as 13% has been attained [1].

Given a discharge current I_d and gas pressure, the hollow-cathode discharge operating voltage can be minimized by the optimization of the electrode geometry. If the anode-to-cathode surface area ratio S_a/S_c meets the condition $S_a/S_c \sim (m/M)^{1/2}$, where m and M are the masses of an electron and ion, respectively, fast electrons in the plasma relax (in terms of the energy) to the maximum degree and the electron current closes on the anode without forming a negative space charge layer [2].

The application of a magnetic field allows one to increase the discharge current or decrease the initiation voltage under low gas pressures. In this case, the character of fast electron oscillation in the gap varies in accordance with the electrode system configuration. While in an inverted magnetron fast electrons largely drift in the azimuth direction, in a reflecting (PIG) system the electrons oscillate mainly between the cathodes along the magnetic field. Weak (≈ 1 mT) magnetic fields (where the Larmor radius of a fast electron is comparable to the dimensions of the electrode system (≈ 0.1 m)) used in an inverted cylindrical magnetron provide the uniform current density distribution for ions extracted from the plasma along the magnetic field [3]. However, the ion extraction efficiency in such a system is usually no greater than 10% because of the large cylindrical surface area of the cathode. The configuration of the reflecting system makes it possible to increase the ratio of the emitting surface area of the plasma to the total surface area of two planar reflecting cathodes. In this system, the loss of fast electrons toward the larger cathode can be limited by applying magnetic fields that are stronger than in an inverted magnetron. This causes a substantial radial nonuniformity of the plasma. According to [4], the near-uniform ion emission current density and a high ion extraction efficiency (up to 30%) can be reached in a modified reflecting system with a nonequipotential cathode. However, relatively high magnetic fields (5–10 mT) used in this system limit the operating voltage of the ion source, causing a discharge to be initiated in crossed electric and magnetic fields within the high-voltage gap between the coaxial electrodes. Therefore, searching for ways of improving the efficiency of ion extraction from the glow-discharge plasma under weak magnetic fields seems to be of current interest.

If an electrode system provides the conditions where the angle Θ between the electric field **E** in the cathode sheath of a glow discharge and an external magnetic field **B** is $0 < \Theta < \pi/2$, fast electrons in the cathode sheath acquire a velocity component along the magnetic field toward the plasma surface. When arriving at the emitting surface, fast electrons, which also drift in the azimuth direction due to the transverse velocity component, may increase the emission current density, as takes place in a reflecting system. At the same time, the possibility of controlling the ion emission current density by varying the magnetic field, which is characteristic of an inverted magnetron, persists.

In this work, we study a glow-discharge plasma emitter used in the electrode system of an inverted magnetron with a conic cathode. Our aim was to improve the extraction of ions from the discharge plasma and the energy efficiency of such emitters.

EXPERIMENTAL

In our experiments, the electrode system of an ion source used in [5] (Fig. 1) served as the basic system. The system comprises a hollow cylindrical cathode made of stainless steel whose diameter D and length L are 150 mm and a rod-shaped tungsten anode 4 mm in diameter and 100 mm long placed on the cathode axis. A magnetic field is generated by a solenoid placed outside the grounded case of the ion source. The values of the induction B are measured on the system's axis in the plane of the solenoid. A two-electrode ion–optical device is used to extract ions and cut off secondary electrons. Either of the electrodes has 61 hole (aperture) 12 mm in diameter with an aggregate surface area $S_e = 70$ cm^2. The inner surface area of the cylindrical hollow cathode is $S_c = 1080$ cm^2. A working gas (argon) is delivered directly into the cathode cavity.

To obtain reproducible I–V characteristics, the cathode surface was pretreated by long-term ($\approx$0.5 h) ion sputtering in the discharge until U_d reached the steady-state value. For a minimal gas flow rate $Q \approx 40$ cm^3/min and a minimal induction $B \approx 1$ mT, which are necessary to sustain a discharge current $I_d \approx 0.5$ A, the ignition voltage was $\approx$1 kV. The need for small flow rates Q and weak fields B, which provide the desired electric strength of the accelerating gap of the high-voltage ion source, is also a reason for the poor energy efficiency of these high-voltage ion sources. The emission characteristics of the ion source that were taken for two accelerating voltages are depicted in Fig. 2 (curves 1, 2). The linear run of the curves indicates that the beam loss at the accelerating electrode of the multiaperture ion optics is low. This is supported by the direct measurement of the current in the circuit of this electrode. This current includes not only the loss current of ions accelerated but also the secondary electron current due to ion–electron emission and the ion current from the
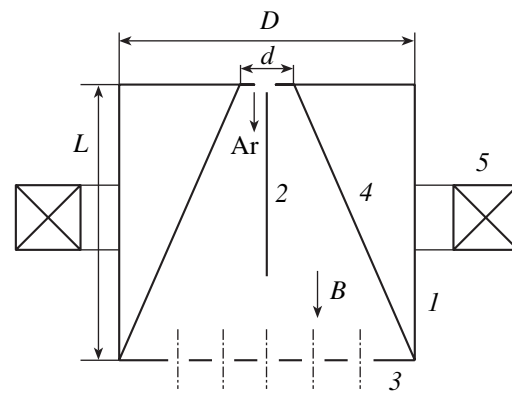

Fig. 1. Electrode system of inverted magnetron type: (*1*) cylindrical hollow cathode, (*2*) rod-shaped anode, (*3*) punched screening electrode of ion optics, (*4*) conic hollow cathode, and (*5*) magnetic coil.
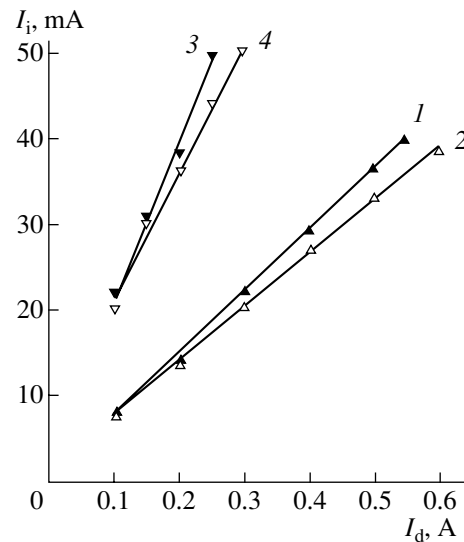
Fig. 2. Emission characteristics of the ion source with (*1*, *2*) cylindrical and (*3*, *4*) conic hollow glow-discharge cathode. The gas flow rate is (*1*, *2*) 30 and (*3*, *4*) 40 cm^3/min. The magnetic induction is $B = $ (*1*, *2*) 2 and (*3*, *4*) 1.5 mT. The accelerating voltage is (*1*, *3*) 30 and (*2*, *4*) 20 kV.

beam plasma. Estimates show that the accuracy of determining the ion beam current I_i from the current in the high-voltage power supply circuit is $\approx$10%. For a cylindrical magnetron, the efficiency α of extracting ions from the plasma that was estimated from the beam-to-discharge current ratio was found to be 7.0–7.5%, which approximately equals the ratio S_e/S_c.

Next, a thin-walled (0.3 mm) truncated cone made of stainless steel was placed into the hollow cathode cylinder (Fig. 1). The diameters of the lower and upper cone bases were $D = 150$ mm and $d = 50$ mm, respectively. The use of the conical cathode with lesser S_c increased substantially the discharge operating voltage ($\approx$1.5 times) at the same values of the flow rate Q and

Fig. 3. I–V characteristics of the conic-hollow-cathode discharge. The gas flow rate is (1–4) 15 and (5, 6) 40 cm³/min. The magnetic induction B is (1, 2) 1, (3, 4) 3, and (5, 6) 0 mT. The accelerating voltage is (1, 3, 5) 0 and (2, 4, 6) 30 kV.
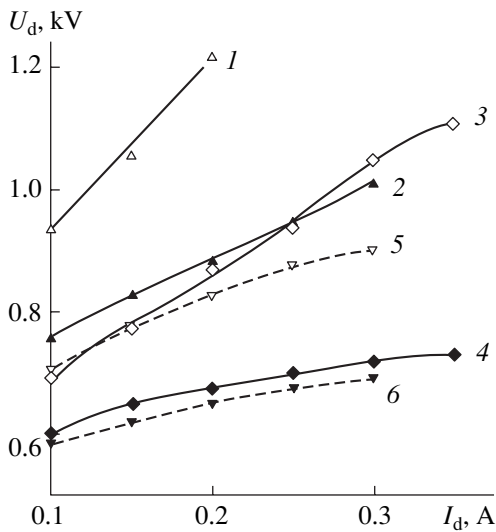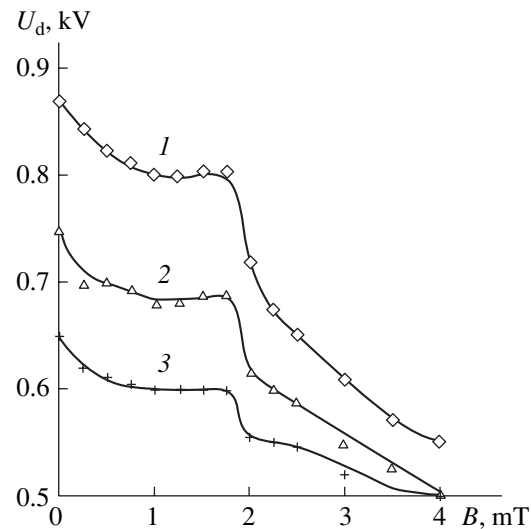
Fig. 4. Discharge operating voltage vs. magnetic field without electron losses through the apertures of the ion optics. The discharge current is 0.2 A. The gas flow rate is (1) 20, (2) 30, and (3) 40 cm³/min.

induction B. Then, we optimized the geometry of the rod-shaped anode. Its length $l = 50$ mm and diameter $d = 1$ mm were experimentally selected so as to minimize the operating values of B_0 and Q_0. After the optimization, the anode-to-cathode surface area ratio S_a/S_c was equal to $\approx 1/300$ (in view of the area of the hole on the cathode end face, through which the anode is introduced into the cathode cavity). This value is close to the ratio $(m/M)^{1/2}$ for argon. Figure 3 (curves 1, 3, 5) shows the I–V characteristics of the discharge for various combinations of B and Q. As a result of the anode geometry optimization, the discharge current in the modified electrode system with the conical cathode was nearly the same, 0.1–0.5 A (for $Q = 20$–40 cm³/min and $B = 0$–4 mT), as in the cylindrical magnetron.

The experiments showed that the I–V characteristics obtained under the ion beam generation conditions differ considerably from those obtained with floating (relative to the Earth potential) electrode potentials in the discharge system. As follows from Fig. 3 (curves 2, 4, 6), the application of a potential to the accelerating gap of the ion optics considerably (by 100–300 V) reduces the value of U_d. As the discharge current increases, the difference between the I–V curves grows. Figure 4 displays the dependences of the discharge operating voltage on the gas flow rate and magnetic field in the conical magnetron operating under the ion collection conditions. When the anode size exceeded the optimal value, the nonmonotonic run of the curves $U_d(B)$ was not observed. From Figs. 3 and 4, one can infer that an increase in the efficiency α of ion extraction from the plasma is the basic way of maintaining a reasonable value of energy efficiency $\eta = \alpha/U_d$ of the ion emitter

used in the high-voltage ion source when the discharge voltage U_d grows sharply with decreasing Q and B.

The emission characteristics of the ion source with the conical cathode are shown in Fig. 2 (curves 3, 4). The ion extraction efficiency in this system rises to 14–21%, with α decreasing with increasing B and increasing with accelerating voltage (Fig. 5).

To estimate the uniformity of ion emission from the plasma in the conical-cathode discharge system, we measured the distribution of the beam current density over the cross section at a distance of 250 mm from the ion source. With the beam current density varying in the range 0.1–0.5 mA/cm² and the accelerating voltage ranging from 10 to 30 kV, the nonuniformity of the distribution is no more than 20% over the central part of the beam 8 cm in diameter. Since the current distribution in a beam of large cross-sectional area also depends on its angular divergence and the distance to the plane of measurement [6], we measured the discharge current distribution in the plane of the screening electrode of the ion optics (this electrode is a part of the cathode). The measurements used planar probes with a working surface area of 0.5 cm² that were under the cathode potential. The results are shown in Fig. 6. As in the cylindrical magnetron, the radial profiles of the current density depend on the magnetic field [3]. For a magnetic field of 1–2 mT, the distribution nonuniformity is within ±5% of the mean current density on a diameter of 9 cm.

DISCUSSION

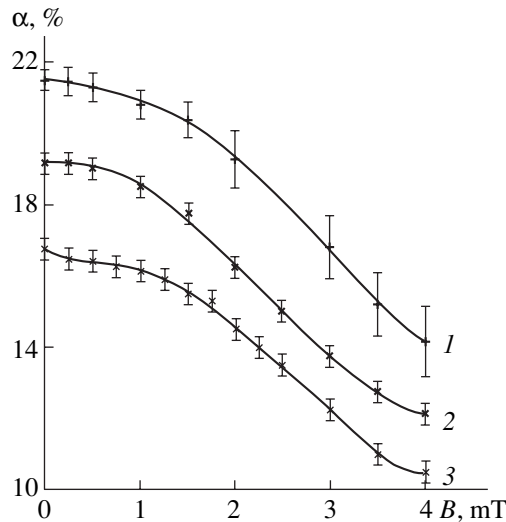Since the total surface area of the conical cathode in the experiments comprises ≈ 0.6 of the surface area of

Fig. 5. Efficiency of ion extraction from the discharge plasma vs. magnetic field. The discharge current is 0.2 A and the gas flow rate is 30 cm^3/min. The accelerating voltage is (*1*) 30, (*2*) 20, and (*3*) 15 kV.
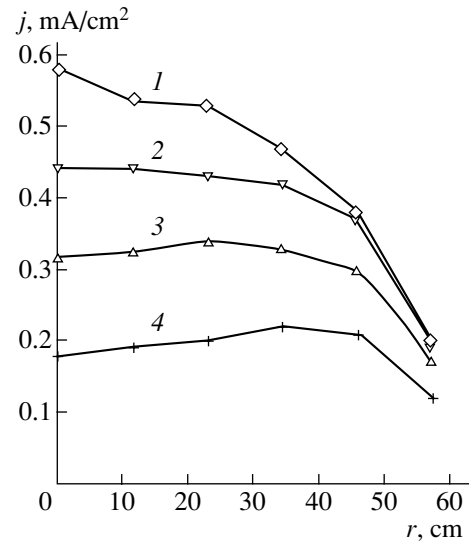
Fig. 6. Discharge current density distribution in the plane of the screening electrode. The discharge current is 0.2 A, and the gas flow rate is 30 cm^3/min. The magnetic induction is (*1*) 0, (*2*) 1, (*3*) 2, and (*4*) 4 mT.

the cylindrical cathode, the corresponding ion extraction efficiency estimated from the ratio S_e/S_c is $\alpha \approx 13\%$; that is, the increment due to the geometrical factor amounts to 5.5–6.0% of the discharge current. The same value, $\alpha = 13\%$, was obtained in a system with a hollow conical cathode without a magnetic field [1]. Hence, it may be argued that the additional contribution (1–8% of the discharge current) to the value of α obtained, $\alpha = 14$–21%, is associated with a change in the electron oscillation character and in the distribution of the discharge current over the cathode. In a conical magnetron, fast electrons moving toward the emitting surface of the plasma enhance the gas ionization rate, causing the ion emission current to grow.

It should be noted that the transfer of fast electrons toward the emitting surface of the plasma also takes place in the absence of a magnetic field. A simple geometrical consideration of electron trajectories shows that each reflection of a fast electron from the cathode sheath on the opposite surface of the cone contributes to its velocity component parallel to the system's axis. After the electron has reflected from the emitting surface, its longitudinal velocity in the backward direction declines. Eventually, because of the anisotropic properties of the discharge, it will concentrate at the screening electrode. Therefore, the ion extraction efficiency in a system with a hollow conic cathode is expected to be greater than the ratio S_e/S_c. However, in [1], such an effect was not confirmed. A high ion extraction efficiency (to 15%) in a conic-hollow-cathode glow discharge system was reported in [7]; however, this result was not related to the effect of the hollow cathode shape and behavior of secondary electrons. The influence of the fast electron reflection conditions in the wedge-

shaped hollow cathode of a plasma electron source on the probability of their return to the cathode and on the fast electron lifetime in the discharge was discussed in [8]. It was assumed that the change in the cathode shape may be a factor that reduces the threshold operating gas pressure in the plasma. The variation of the degree of ion sputtering of the hollow cathode conic surface, which was observed in our experiments, confirms indirectly that the plasma is anisotropic in the axial direction. As the screening electrode was approached, the intensity of cathodic sputtering grew.

In a cylindrical magnetron, the radial distribution of the ion emission current density is controlled by varying the Larmor radius of fast electrons, which drift in the azimuth direction, i.e., by varying the effective range of ionization. As a result of the field enhancement, the current density distribution with a near-axis peak transforms first into a plane distribution and then into a distribution with a dip at the axis. In the system suggested, the drift of fast electrons in a weak magnetic field persist, the oscillating electron trajectories at the base of the cone being almost the same as in a cylindrical magnetron. The decrease in the Larmor radius of fast electrons $\rho = (1/B)(m/e)V_\perp$ due to a decrease in the transverse (relative to the direction of the field **B**) component $V_\perp$ of the electron velocity is bound to somewhat reduce the optimal value of B compared with a cylindrical magnetron. However, possible changes are within the accuracy of the current density profile measuring technique.

The high discharge operating voltage inherent in the electrode system of a plasma ion emitter is due to great plasma particle losses in apertures in the screening electrode of the ion–optical system. This makes it diffi-

cult to sustain the discharge. The geometry of the multiaperture ion–optical system used in the experiments was optimized so as to diminish beam current losses at the accelerating electrode when the beam diverges because of its self-space charge. By means of mathematical simulation and experimental investigation, we realized beam formation conditions such that the crossovers of elementary beams (i.e., beams that were generated in a separate aperture), 1–3 mm in size, were formed in apertures of the accelerating electrode, which had a much larger diameter (12 mm). Therefore, a displacement of the elementary beam trajectory did not cause any appreciable losses in a wide range of beam parameters. However, when the apertures in the screening electrode are too wide, an increase in the discharge current may break the cathode sheath near the apertures. Under ambipolar diffusion conditions, particles of the plasma issuing from the apertures move to the walls and recombine without sustaining the discharge. When a high voltage is applied, fast electrons are reflected by the field of the accelerating electrode, turn back to the plasma, and spend a part of their energy on gas ionization. As a result, the discharge operating voltage decreases (Fig. 3). As the discharge current grows, so do electron losses, since, to a first approximation, they depend on the relationship between the ion optics aperture size and cathode sheath thickness.

For a singly charged argon ion current density of 0.5 mA/cm^2 and a voltage drop across the sheath of 700 V, the cathode sheath thickness estimated in the Child–Langmuir approximation was found to be $l \approx 0.55$ cm. This means that the sheath may break. In addition, one should take into account that the discharge operating voltage is influenced mainly by the loss of fast electrons, which may escape without breaking the cathode sheath through the near-axis part of the apertures, where the potential barrier lowers due to the electric field in the accelerating gap. In the conical magnetron, an increase in the voltage because of the loss in fast electrons was several times greater than in the cylindrical device.

The elevated discharge operating voltage under the ion extraction conditions provides the constancy of the discharge current. The effect of ion losses on the discharge operating voltage may be estimated with the results of [9], from which it follows that the initiation voltage rises in proportion to $1/(1 - \alpha)$ provided that the electrons in the plasma completely relax. Thus, the increment of the initiation voltage due to ion losses is $\approx 25\%$ (150–200 V) for $\alpha \approx 20\%$ obtained in our experiments. For the cylindrical magnetron with $\alpha \approx 7.5\%$, the increment is only $\approx 8\%$.

The efficiency α depends on the accelerating voltage because of the variable curvature of the plasma meniscus in the ion optics aperture. An increase in the surface area from which electrons enter into the accelerating gap causes α to grow with accelerating voltage [10]. In our experiments, the effect of the magnetic field

on the discharge operating voltage appears as a result of increasing the ratio between the path length of an oscillating electron in the cathode sheath and the path length of an electron in the plasma, $\sim l/2\rho$, rather than as a result of the reliable confinement of fast electrons. This increases the ionization probability in the sheath and the probability of producing secondary electrons with an energy sufficient for gas ionization. Such a mechanism of generating secondary electrons is energetically more favorable than ion–electron emission with the subsequent accelerating of electrons in the cathode sheath [11, 12]. Therefore, the discharge operating voltage declines rapidly with decreasing Larmor radius of fast electrons. However, as the field B grows, the average number of reflections of an electron from the conical surface increases on its way to the emitting surface of the plasma and the Larmor radius of fast electrons diminishes. As a result, a fraction of the discharge current accounted for by the lateral conical surface increases and the efficiency of ion extraction from the plasma drops.

For the same values of the gas flow rate Q and induction B, the energy efficiency $\eta = I_i/U_d I_d = \alpha/U_d$ of the ion source with the modified electrode system was found to be 2.0–2.5 times higher than in the cylindrical magnetron. The absolute values of η for the conic-cathode source are 0.20–0.25 A/kW for $I_d = 0.1$–0.3 A for the moderate values of magnetic induction and gas flow rate used in our experiments.

CONCLUSIONS

The modification of the hollow cathode geometry in an inverted magnetron allowed us to raise the energy efficiency of the ion emitter 2.0–2.5 times. At the same time, the near-uniform distribution of the emission current density and moderate values of the magnetic induction ($B = 1$–2 mT) and gas flow rate ($Q = 20$–30 cm^3/min), which are favorable to achieving high (several tens of kilovolts) ion source operating voltages, are retained.

These results reflect a change in the oscillations of electrons emitted by the cathode and accelerated in the cathode sheath. The formation of a fast electron flow toward the emitting surface of the plasma increases the ion-emitting plasma density.

The high discharge operating voltage in the electrode system of the effective glow-discharge ion emitter is explained by high losses of charged plasma particles through the apertures in the ion optics electrodes. The contribution of the electron losses to the increment of the discharge operating voltage depends on the ratio between the aperture size and cathode sheath thickness.

REFERENCES

1. V. N. Glazunov, V. G. Grechanyi, and A. S. Metel', Prib. Tekh. Éksp., No. 1, 145 (1988).
2. A. S. Metel', Zh. Tekh. Fiz. **54**, 241 (1984) [Sov. Phys. Tech. Phys. **29**, 141 (1984)].
3. N. V. Gavrilov, G. A. Mesyats, S. P. Nikulin, *et al.*, J. Vac. Sci. Technol. A **14**, 1050 (1996).
4. S. P. Nikulin and D. F. Chichigin, in *Abstracts of the Conference on the Physics of Low-Temperature Plasma (FNTP-2001), Petrozavodsk, 2001*, Part 1, pp. 248–252.
5. N. V. Gavrilov, S. P. Nikulin, and G. V. Radkovskiǐ, Prib. Tekh. Éksp., No. 1, 93 (1996).
6. S. I. Molokovskiǐ and A. D. Sushkov, *Intense Electron and Ion Beams* (Énergoatomizdat, Moscow, 1991).
7. S. M. Chesnokov, USSR Inventor's Certificate No. 1598757, 1989.
8. V. Ya. Martens, Zh. Tekh. Fiz. **69** (7), 135 (1999) [Tech. Phys. **44**, 860 (1999)].
9. S. P. Nikulin, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 9, 63 (2001).
10. N. V. Gavrilov and D. Emlin, Zh. Tekh. Fiz. **70** (5), 74 (2000) [Tech. Phys. **45**, 597 (2000)].
11. V. N. Glazunov, V. G. Grechanyǐ, and A. S. Metel', Zh. Tekh. Fiz. **52**, 1767 (1982) [Sov. Phys. Tech. Phys. **27**, 1084 (1982)].
12. V. A. Nikitinskiǐ and B. I. Zhuravlev, Zh. Tekh. Fiz. **52**, 880 (1982) [Sov. Phys. Tech. Phys. **27**, 563 (1982)].

Translated by V. Isaakyan

Breakdown Delay in a Vacuum

A. A. Emel'yanov

Orel State Technical University, Orel, 302020 Russia
e-mail: orelrce@ostu.ru
Received October 7, 2002

Abstract—Breakdown delay in a vacuum is considered in terms of the Joule mechanism. The effect of the cathode material on the delay time is studied. A test for optimality for the gap electrode surface condition is stated. When prepared under optimal conditions, cathodes have a minimum field enhancement coefficient at surface microirregularities. This allows one to estimate the emission parameters of the cathode surface and the dielectric strength of vacuum insulation. © 2003 MAIK "Nauka/Interperiodica".

INTRODUCTION

A high-voltage pulse applied to the electrodes of a vacuum gap may result in electric breakdown, which somewhat lags behind the pulse. A breakdown delay in a vacuum is determined by the inertia of processes initiating breakdown. The delay time t_d depends on the surface condition, the electrophysical parameters of the cathode material, and the amplitude and shape of the pulse applied.

The use of nanosecond voltage pulses cuts considerably the number of breakdown-initiating mechanisms. The cathodic mechanism, where breakdown is initiated by the explosive destruction of the emitter that is Joule-heated by passing thermionic and field-emission currents, is the main one in this case.

For the Joule mechanism, there are analytical relations between the breakdown delay time in a vacuum, electric field intensity, and physical constants of the cathode material. These relations allow one not only to explain experimental results but also to analyze the effect of the electrode material on the breakdown delay time and also to formulate a test for optimality for electrode surface preparation conditions. The treatment of the electrodes by high-voltage pulses of duration t_p equal to the breakdown delay time, $t_p = t_d$, corresponds to the optimal preparation conditions. Such a regime of cathode surface preparation minimizes the electric field enhancement coefficient β at surface microirregularities and maximizes the pulsed insulation strength. In addition, the optimal preparation conditions allow one to reach a desired vacuum insulation strength and estimate the emission parameters of the cathode surface [1].

TEST FOR INITIATION

The use of voltage pulses with a duration

$$t_r \ll t_p \ll h^2 \rho c/\lambda, \tag{1}$$

where t_r is the time of thermal relaxation; h is the emitter height; and ρ, c, and λ are the density, specific heat, and thermal conductivity of the emitter, respectively, reduces the boundary-value problem for the emitter temperature distribution [2] to the form

$$\begin{cases} \rho c \dfrac{dT}{dt} = j_0^2 \left(\dfrac{\pi T/2T^*}{\sin(\pi T/2T^*)} \right)^2 \kappa_0 T, \\ T|_{t=0} = T_0, \quad T|_{t=t_d} = T_{cr}. \end{cases} \tag{2}$$

Here, j_0 is the field emission current density; κ_0 is the proportionality factor in the temperature dependence of the emitter resistivity, $\kappa(T) = \kappa_0 T$; T^* is the inversion temperature, $T^* = 5.67 \times 10^{-7} \varphi^{-1/2}$; and φ is the work function [3].

The solution of Eq. (2) yields analytical expressions for the delay time t_d as a function of the critical field intensity E_{cr} at an emitting microarea (hereafter microintensity) and physical constants of the cathode material without going into the geometry of this area.

Written in integral form,

$$\int_0^{t_d} j_0^2(t) dt = a \frac{\rho c}{\kappa_0}, \tag{3}$$

where

$$a = \left[Ci(x) - \frac{\sin x}{x} - \frac{1}{2} \frac{\sin^2(x/2)}{(x/2)^2} \right] \Bigg|_{x = \pi T_0/T^*}^{x = \pi T_{cr}/T^*},$$

$$Ci(x) = -\int_x^\infty \frac{\cos y}{y} dy.$$

Equation (2) is a test for vacuum breakdown initiation. According to this test, breakdown occurs when the energy evolution in the emitter becomes equal to the energy of emitter destruction, causing explosion [4].

Physical properties of metals

Metal	ρ, kg/m³	T_{le}, °C	λ, W/(m °C)	c, J/(kg °C)	k, 10^{-10}, (Ω m)/K	ϕ, eV	$\rho c / k_0$, 10^{16} (A² s)/m⁴
Zr	6500	1860	17	276	15.02	3.70	0.12
Ti	4500	1725	15	577	17.58	3.85	0.15
Be	1850	1284	167	200	1.46	3.90	0.25
Nb	8570	2410	50	272	5.13	4.00	0.45
Ta	16700	2850	54	142	4.94	4.10	0.48
Al	2703	657	209	922	0.96	4.25	2.60
Mo	10200	2620	151	264	1.90	4.27	1.42
Ag	10500	961	408	234	0.59	4.30	4.16
Au	19300	1063	311	126	0.81	4.30	3.00
Fe	7870	1535	73	452	3.59	4.31	1.00
Cu	8890	1083	394	386	0.63	4.40	5.44
Ni	89001	1455	95	444	2.67	4.50	1.48
W	19300	3380	168	218	2.01	4.54	2.09
Cr	7100	1890	67	461	7.69	4.58	0.43
Re	20500	3180	71	138	7.69	4.80	0.37

The value of $a\rho c / \kappa_0$ is the specific energy of emitter destruction.

BREAKDOWN DELAY TIME

Experimentally, the delay time t_d is determined by using rectangular voltage pulses with different ratios of the leading edge time t_{le} and pulse duration t_p. Note that the values of t_d determined by using rectangular pulses with an infinitely short leading edge time ($t_{\text{le}} \ll t_p$) differ both from t_{d1} measured using rectangular pulses with a finite leading edge time ($t_{\text{le}} \sim t_p$) and from t_{d2} determined at the pulse front ($t_{\text{le}} = t_p$).

The solution of Eq. (3) yields expressions for the delay time for pulses of different shapes:

for rectangular pulses with an infinitely short leading edge time,

$$t_d = 4.2 \times 10^{11} \frac{a\rho c}{\kappa_0} \phi^2 \exp\{-21.6\phi^{-1/2}\}$$
$$\times \frac{\exp\{1.32 \times 10^{10} \phi^{3/2} E_{\text{cr}}^{-1}\}}{E_{\text{cr}}^4}; \quad (4)$$

for skewed pulses,

$$t_{d2} = 1.32 \times 10^{10} \frac{\phi^{3/2}}{E_{\text{cr}}} t_d; \quad (5)$$

and for rectangular pulses with a finite leading edge time,

$$t_{d1} = t_d + t_{\text{le}}(1 - 7.58 \times 10^{-11} \phi^{-3/2} E_{\text{cr}}). \quad (6)$$

The dependences $t_d = f(E_{\text{cr}})$ following from Eqs. (4)–(6) for tungsten are in satisfactory agreement with the experimental results [5], which substantiates the approach used and the assumptions made. With $E_{\text{cr}} = $ const, the delay time t_{d2} (skewed wave) is one order of magnitude longer than the time t_d (rectangular pulse), which is in agreement with experimental data [6].

EFFECT OF THE ELECTRODE MATERIAL

Analytical expressions (4)–(6) imply that the breakdown delay time is a function of the critical electric field microintensity, work function, and specific energy of emitter destruction. Equation (4) was used to construct the dependences $t_d = f(E_{\text{cr}})$ for metals with a work function ranging from 3.7 to 4.8 eV. The physical constants [7–9] used in the calculations are listed in the table.

The breakdown delay time as a function of the electric field microintensity initiating breakdown is shown in Fig. 1 for zirconium, beryllium, tantalum, molybdenum, and copper. The curves were calculated according to Eq. (4).

With $E_{\text{cr}} = $ const, the delay time for copper is four orders of magnitude larger than for zirconium. The dependences $t_d = f(E_{\text{cr}})$ calculated for nickel, tungsten, chromium, and rhenium differ slightly from the curve for copper, almost coinciding with it. It is seen that the work function ϕ of the material and the specific energy of emitter destruction $a\rho c / \kappa_0$ affect significantly the delay time t_d.

The specific energy of destruction $a\rho c / \kappa_0$ for different metals as a function of the work function is shown in Fig. 2a.
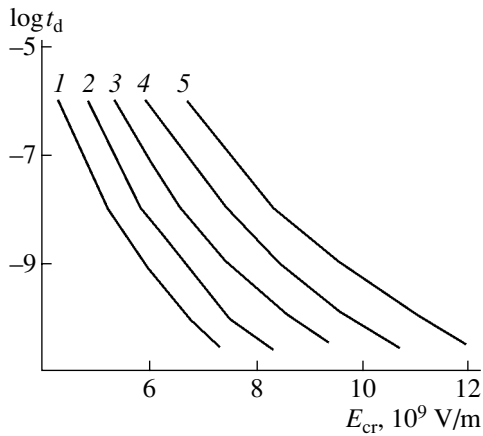
Fig. 1. Delay time vs. the electric field microintensity for different metals: (*1*) Zr, (*2*) Be, (*3*) Ta, (*4*) Mo, and (*5*) Cu. t_d is given in seconds.
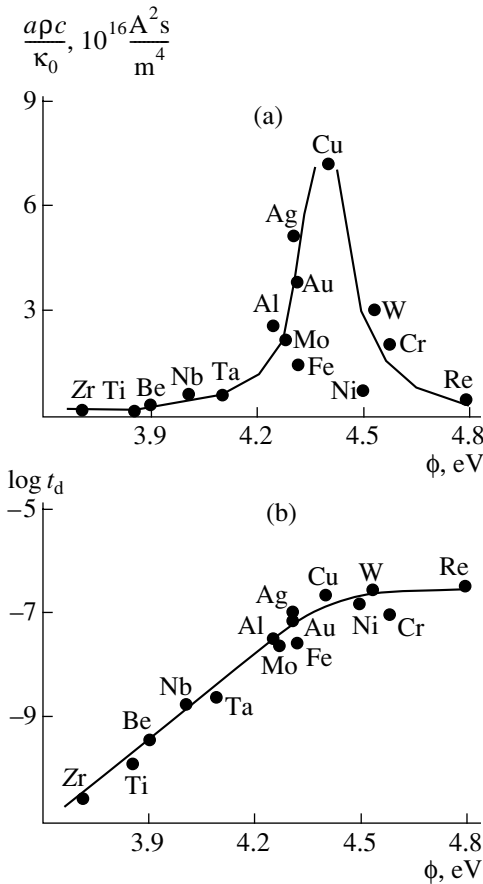

Fig. 2. (a) Specific energy of emitter destruction and (b) the delay time of breakdown vs. the work function. t_d is in seconds.

The transition from zirconium ($\varphi_{Zr} = 3.7$ eV) to copper ($\varphi_{Cu} = 4.4$ eV) increases the work function by 0.7 eV and the specific energy of destruction roughly 40 times. The latter parameter is maximum for copper, $a\rho c/\kappa_0 = 7.24 \times 10^{16}$ A^{2}s/m^4. The transition from copper to rhenium ($\varphi_{Re} = 4.8$ eV) increases φ by 0.4 eV; however, $a\rho c/\kappa_0$ decreases ≈ 13 times to 5.4×10^{15} A^{2}s/m^4. The presence of the maximum in the dependence $a\rho c/\kappa_0 = f(\varphi)$ substantially affects the behavior of the curve $t_d = f(\varphi)$.

The logarithm of the time delay vs. the work function is plotted in Fig. 2b for different metals at $E_{cr} = 7 \times 10^9$ V/m.

The transition from zirconium to copper increases the delay time by four orders of magnitude. Here, two orders are due to an increase in the work function and the other two orders are due to an increase in the specific energy of destruction. The transition from copper to rhenium almost does not change the delay time, since the growth of t_d due to an increase in φ is compensated for by the corresponding decrease in $a\rho c/\kappa_0$.

The results presented allow us to substantiate the choice of the electrode material for vacuum gaps. To provide a desired dielectric strength, preference must be given to materials with a high work function and a high specific energy of destruction.

TEST FOR OPTIMALITY

If the high-voltage pulse duration is less than the breakdown delay time, $t_p < t_d$, the pulse energy cannot destroy the emitter. As the duration increases and the critical breakdown-initiating state, which is characterized by the equality $t_p = t_d$, is approached, the surface microrelief smoothes and the polishing of the cathode surface starts [10].

For $t_p > t_d$, the high-voltage pulse energy is spent not only on emitter destruction but also on switching processes in the gap. As the duration of the explosive emission current pulse decreases and simultaneously the high-voltage pulse duration approaches the critical value, which is equal to the breakdown delay time in a vacuum ($t_p = t_d$), microcraters shrink and the cathode surface becomes smooth (nearly polished) [11, 12].

For $t_p = t_d$, the energy stored in the emitter before explosion becomes equal to the energy of emitter destruction. The pulse energy here is sufficient for breakdown initiation but falls short of sustaining and extending the discharge. The treatment of the cathode surface by high-voltage pulses with a duration $t_p = t_d$ only destroys microasperities, but new emission centers do not emerge.

Thus, the optimal regime of treatment of the vacuum gap electrodes is that when the duration of a high-voltage pulse is equal to the breakdown delay time, $t_p = t_d$. In this case, the cathode surface becomes smooth and offers the maximum pulsed dielectrical strength at a minimal duration of surface conditioning. A test for

optimality is obtained by substituting $t_p = t_d$ into Eq. (3):

$$\int_0^{t_p} j_0^2(t)dt = a\frac{\rho c}{\kappa_0}. \tag{7}$$

If the pulse power grows while the pulse shortens so that the equality $t_p = t_d$ remains valid, the efficiency of electrode conditioning is improved and finally the cathode surface is polished.

ESTIMATION OF EMISSION PARAMETERS

In experiments on breakdown delay determination, the electrodes of a vacuum gap are pretrained by high-voltage pulses (t_p = const) with a small overvoltage to make every value of the conditioning pulse amplitude correspond to a stable value of the delay time and, hence, the stable state of the cathode surface. The stable state of the cathode surface is described by the coefficient of electric field enhancement β at surface microirregularities. The local field and the macrofield are related as

$$E = \beta E_0. \tag{8}$$

The experimental dependence $t_d = f(E_0)$ obtained as a result of high-voltage training with a minimum overvoltage reflects the optimal surface conditioning regime ($t_p \approx t_d$). The enhancement coefficient β under the optimal conditioning regime can be determined by comparing the experimental curve $t_d = f(E_0)$ and calculated curve $t_d = f(E_{cr})$ with t_d = const. The comparison of the experimental curves $t_d = f(E_{cr})$ for pulses of different shapes with E_0 = const allows one to estimate both the field enhancement coefficient and the work function.

(1) Field enhancement coefficient. Experimental data for breakdown delay times in a vacuum are most fully presented for copper electrodes. They have been determined in an operating and ultrahigh vacuum over a wide range of electrode spacings from 3 μm to 20 cm (i.e., spanning five orders of magnitude) [13–17]. The experimental, $t_d = f(E_0)$ (curve 1), and calculated, $t_d = f(E_{cr})$ (curve 2), dependences of the vacuum breakdown delay time on the breakdown-initiating electric field intensity are shown in Fig. 3 for copper electrodes. The calculated dependence $t_d = f(E_{cr})$ is constructed by using Eq. (4) and the physical constants given in the table.

The difference between the abscissas of the experimental, $t_d = f(E_{cr})$, and calculated, $t_d = f(E_0)$, curves at t_d = const gives the field enhancement coefficient

$$\log\beta = \log E_{cr} - \log E_0, \tag{9}$$

which is reached under the optimal conditioning regime.

As follows from the plots, as the electric field intensity increases, curves $t_d = f(E_0)$ and $t_d = f(E_{cr})$ approach
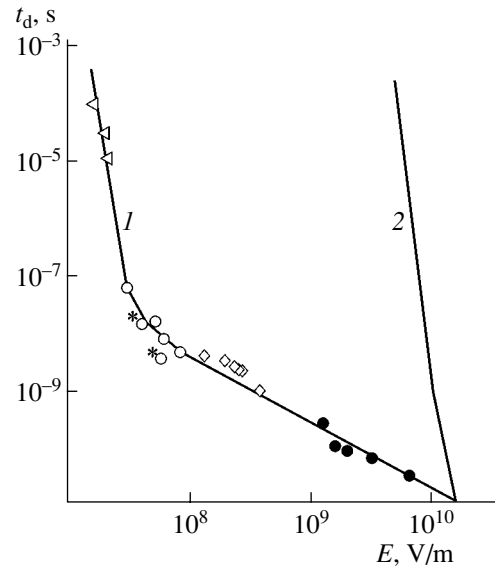

Fig. 3. Breakdown delay time vs. the breakdown field intensity for copper electrodes. (1) Experimental dependence $t_d = f(E_0)$ and (2) calculated dependence $t_d = f(E_{cr})$. Experimental data from (○) [13], (◁) [14], (□) [15], (*) [16], and (●) [17].

each other. They coincide at a certain critical value

$$E_0 = E_{cr} = 1.32 \times 10^{10} \text{ V/m}, \tag{10}$$

when the intensities at the top and base of a microasperity become the same and the field enhancement coefficient reaches the minimum value $\beta = 1$, which corresponds to the ideally smooth surface.

Comparing the experimental and calculated curves at t_d = const, the dependences of the field enhancement coefficient $\beta = f(E_0)$ and $\beta = f(t_p)$ on the surface conditioning parameters can be constructed. The curves $\beta = f(E_0)$ and $\beta = f(t_p)$ (Fig. 4) describe the variation of the coefficient β when the surface was prepared under optimal conditions ($t_p = t_d$) for different values of macrointensity E_0.

The field enhancement coefficient β obtained at $t_p = t_d$ is approximated by the power relationship [18]

$$\beta = 1.28 \times 10^9 E_0^{-0.9}. \tag{11}$$

According to (11), the cathode surface prepared under the optimal conditions has a desired value of the coefficient β, which depends only on the breakdown field macrointensity E_0.

Under the optimal preparation conditions, the field enhancement coefficient β may vary by three orders of magnitude depending on the amplitude and duration of high-voltage pulses. This could be used to produce a microrelief with a desired value of β.

A kink in the curve $\beta = f(t_p)$ shown in Fig. 4b at $t_d \approx 10^{-7}$ s is apparently related to the effect of emitter cool-
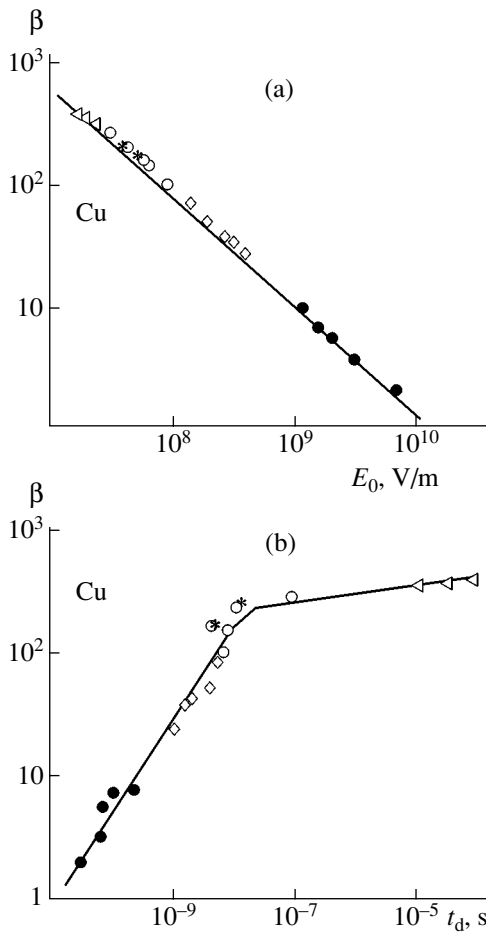
Fig. 4. Field enhancement coefficient vs. the conditioning parameters for $t_p \approx t_d$. (a) $\beta = f(E_0)$ and (b) $\beta = f(t_p)$. Experimental results from (○) [13], (◁) [14], (□) [15], (*) [16], and (●) [17].

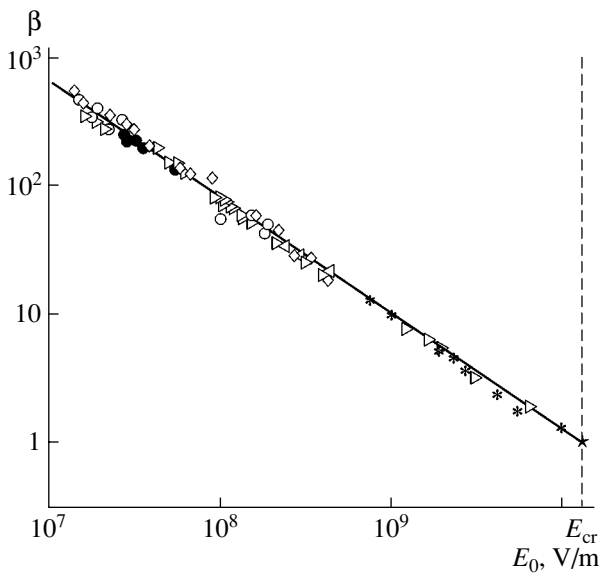**Fig. 5.** Field enhancement coefficient after surface conditioning by pulses with $t_p = t_d$ vs. the macrointensity of the breakdown field: (◇) Al [13–16], (●) Cr [19], (○) Fe [6, 18], (*) Ni [17], (▷) Cu [13–17], (◁) Mo [15, 16], and (*) W [6, 17].

ing by heat conduction. It follows from this curve that, to improve the dielectric strength of vacuum insulation, it is appropriate to apply voltage pulses of duration $t_p < 10^{-7}$ s. Note that the conditioning efficiency increases with decreasing pulse duration.

The field enhancement coefficient β, which is defined as the ratio of the intensities at the top and base of a microasperity, depends only on its geometry and is independent of the electrode material. Hence, expressions (10) and (11), which were derived for copper electrodes, must also be valid for other materials. To verify this assumption, the calculations for aluminum, chromium, iron, nickel, molybdenum, and tungsten electrodes were carried out. The experimental dependences were constructed with data reported elsewhere [5, 6, 13–17, 19, 20].

The field enhancement coefficient β determined in the vacuum breakdown delay experiments performed under the optimal high-voltage pulse conditioning ($t_p \approx t_d$) are plotted against E_0 in Fig. 5 for aluminum, chromium, iron, nickel, copper, molybdenum, and tungsten electrodes.

The experimental data for the electrodes made of different materials fall on a single line $\beta = f(E_0)$. It follows from this dependence that the coefficient β, which characterizes the state of the cathode surface treated by high-voltage pulses of duration $t_p = t_d$ is independent of the electrode material and varies only with the breakdown-initiating electric field macrointensity E_0.

With test for optimality (7) fulfilled, an increase in the macrointensity E_0 leads to a decrease in the enhancement coefficient β; that is, the cathode surface quality is improved. For $E_0 = E_{cr}$ and $t_p = t_d$, the pulsed treatment polishes the cathode surface.

From expressions (5) and (6) for the delay time in the case of a skewed voltage wave and rectangular pulse with a finite leading-edge time, respectively, and taking into account expression (8), we obtain analytical estimates of the field enhancement coefficient β, using the experimentally found breakdown delay time for a skewed voltage wave,

$$\beta = 1.32 \times 10^{10} \varphi^{3/2} E_0^{-1} \frac{t_d}{t_{d2}}, \qquad (12)$$

and a rectangular pulse with a finite leading-edge time,

$$\beta = 1.32 \times 10^{10} \varphi^{3/2} E_0^{-1} \left(1 + \frac{t_d - t_{d1}}{t_{le}}\right). \qquad (13)$$

From Eqs. (12) and (13), it follows that the coefficient β can be found by comparing breakdown delay times for pulses of different shape with $E_0 = \text{const}$. An experimental check with copper electrodes showed that, with $E_0 = 4.2 \times 10^7$ V/m, the delay time for the rectangular pulse is $t_d = 28$ ns and for the skewed wave $t_{d2} = 0.5$ ns. The value of β determined by Eq. (12) was found to be 162, which corresponds to the condition

$t_p = t_d$ with $E_0 = 4.2 \times 10^7$ V/m and agrees with the dependence $\beta = f(E_0)$ shown in Fig. 5.

(2) Work function. The cathode surface prepared under optimal conditions has a known coefficient of the electric field enhancement β, which allows us to estimate the work function. Solving Eqs. (12) and (13) for φ gives

$$\varphi = 0.21 E_0^{1/15}(t_{d2}/t_d)^{2/3}, \tag{14}$$

$$\varphi = 0.21 E_0^{1/15}(t_{le}/(t_d - t_{d1} + t_{le}))^{2/3}. \tag{15}$$

With $E_0 = $ const, we can now estimate the work function from experimental breakdown delay times for voltage pulses of various shape.

The tests were carried out with skewed and rectangular pulses applied to aluminum and copper electrodes. In the former case, the breakdown time delay with $E_0 = 1.9 \times 10^7$ V/m was equal to $t_{d2} = 1.3 \times 10^{-5}$ s for the skewed wave, while for the rectangular pulse, it was one order of magnitude smaller: $t_{d2} = 7.4 \times 10^{-7}$ s. Based on Eq. (14) and the test results, we estimated the work function for aluminum as $\varphi_{Al} = 4.4$ eV. The measurements for copper electrodes gave $\varphi_{Cu} = 4.7$ eV. These values of the work function are in satisfactory agreement with those listed in the table.

DIELECTRIC STRENGTH

Surface conditioning by pulses with $t_p = t_d$ forms a microrelief characterized by a certain value of the coefficient β (Eq. (11)), which makes it possible to write the delay time as a function of the electric field microintensity E_0. For a rectangular pulse with an infinitely short leading-edge time, we have

$$t_d = 1.57 \times 10^{-25} \frac{a\rho c}{\kappa_0} \varphi^2 \exp\{-21.6\varphi^{-1/2}\}$$
$$\times \frac{\exp\{10.3\varphi^{3/2} E_0^{-0.1}\}}{E_0^{0.4}}; \tag{16}$$

for a skewed wave,

$$t_{d2} = 10.3 \frac{\varphi^{3/2}}{E_0^{0.1}} t_d; \tag{17}$$

and for a rectangular pulse with a finite leading-edge time,

$$t_{d1} = t_d + t_{le}(1 - 0.097\varphi^{-3/2} E_0^{0.1}). \tag{18}$$

Expressions (16)–(18) allow one to calculate the delay time under the optimal conditions for cathode surface treatment (high-voltage pulses with $t_p = t_d$) as a function of the electric field microintensity E_0 if the physical parameters of the cathode material are known.

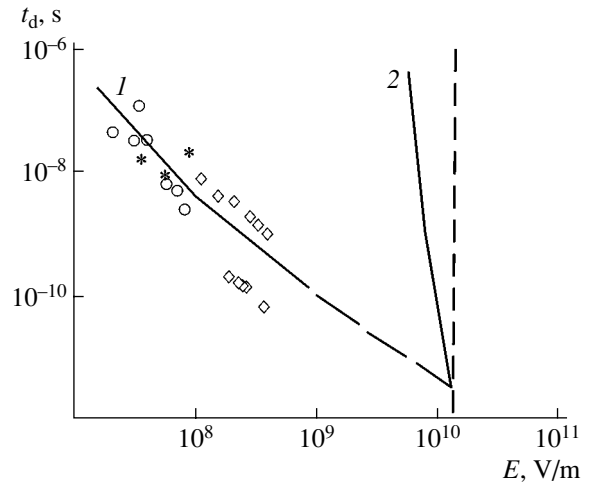The dependences of the delay time on the macrointensity, $t_d = f(E_0)$, and microintensity, $t_d = f(E_{cr})$, of the

Fig. 6. Calculated dependences of the delay time on the breakdown field intensity for aluminum. (*1*) $t_d = f(E_0)$ and (*2*) $t_d = f(E_{cr})$. Experimental results from (○) [13], (◇) [15], and (*) [16].

electric field for aluminum that were calculated based on Eqs. (4) and (16) and the physical parameters listed in the table are shown in Fig. 6. The calculations are in satisfactory agreement with the experimental results [13, 15, 16].

To see whether the Joule mechanism of breakdown can be used to evaluate the pulsed electric strength of a vacuum gap with evaporated electrodes, we performed experiments with an accelerating gap between a microchannel plate and a cathodoluminescent screen.

The dependences of the delay time on the macrointensity, $t_d = f(E_0)$, and microintensity, $t_d = f(E_{cr})$, of the electric field for chromium were calculated based on Eqs. (4) and (16) and the physical parameters (see table). The experimental results for the breakdown delay time in the "microchannel plate–cathodoluminescent screen" evaporated electrode system are in satisfactory agreement with the calculated curve $t_d = f(E_0)$ [20].

The experimental results for the breakdown delay time in the vacuum gaps with both the all-metal and evaporated electrodes show that the coefficient β of electric field enhancement that is observed under the optimal conditioning regime is insensitive to the electrode material and structure and depends only on the breakdown field macrointensity (Eq. (11)).

CONCLUSIONS

Experimental investigation of the breakdown delay in a vacuum provides information about the pulsed dielectric strength of a vacuum gap and the emission parameters of the cathode surface. The treatment of the vacuum gap electrodes by high-voltage pulses with a duration equal to the breakdown delay time ensures an optimal surface conditioning regime. This regime pro-

vides the maximum smoothness of the cathode surface and, accordingly, the highest possible pulsed dielectric strength of vacuum insulation. The cathode surface with a desired coefficient of electric field enhancement β (i.e., vacuum gaps with a desired pulsed dielectric strength) can be obtained by controlling the conditioning pulse power in the optimal regime. On the cathode surface formed in the optimal conditioning regime, one can evaluate both the field enhancement coefficient at surface microirregularities and the work function.

REFERENCES

1. A. A. Emel'yanov, *Pulsed Electrical Treatment of Electrodes in Vacuum* (Ust'-Kamenogorsk, 1999).

2. E. A. Litvinov, G. A. Mesyats, and A. F. Shubin, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 4, 149 (1970).

3. A. A. Emel'yanov, G. M. Kassirov, and G. V. Smirnov, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 4, 142 (1976).

4. A. A. Emel'yanov and G. M. Kassirov, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 9, 105 (1976).

5. G. K. Kartsev, G. A. Mesyats, D. I. Proskurovskiĭ, *et al.*, Dokl. Akad. Nauk SSSR **192**, 309 (1970) [Sov. Phys. Dokl. **15**, 475 (1970)].

6. N. F. Olendzskaya and M. A. Sal'man, Zh. Tekh. Fiz. **40**, 333 (1970) [Sov. Phys. Tech. Phys. **15**, 242 (1970)].

7. *Dictionary of Physics*, Ed. by B. A. Vvedenskiĭ and B. M. Vul (Sov. Éntsiklopediya, Moscow, 1966), Vol. 5.

8. *Physics Encyclopedia*, Ed. by A. M. Prokhorov (Sov. Éntsiklopediya, Moscow, 1983).

9. N. P. Bogoroditskiĭ, V. V. Pasynkov, and B. M. Tareev, *Materials for Electrical Engineering* (Énergoatomizdat, Leningrad, 1977).

10. B. Juttner, V. F. Puchkarov, and W. Rohrbeck, *Nanosecond Field Emission: Production and Destruction of Field Emitting Microtips by Cathode Flares* (Akad. Wiss., Berlin, 1975).

11. G. A. Mesyats, D. I. Proskourovsky, and E. B. Yankelevitch, in *Proceedings of the 7th International Symposium on Discharges and Electrical Insulation in Vacuum (ISDEIV), Novosibirsk, 1976*, pp. 230–233.

12. G. A. Mesyats, D. I. Proskourovsky, E. B. Yankelevitch, *et al.*, Dokl. Akad. Nauk SSSR **277**, 1335 (1976) [Sov. Phys. Dokl. **21**, 228 (1976)].

13. G. M. Kassirov, Zh. Tekh. Fiz. **36**, 1883 (1966) [Sov. Phys. Tech. Phys. **11**, 1403 (1966)].

14. I. I. Kalyatskiĭ, G. M. Kassirov, G. V. Smirnov, *et al.*, Zh. Tekh. Fiz. **45**, 1547 (1975) [Sov. Phys. Tech. Phys. **20**, 988 (1975)].

15. G. A. Mesyats, S. P. Bugaev, D. I. Proskourovsky, *et al.*, Radiotekh. Élektron. (Moscow) **14**, 2222 (1969).

16. S. P. Vavilov and G. A. Mesyats, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 8, 90 (1970).

17. B. Juttner, W. Rohrbeck, and H. Wolff, in *Proceedings of the IX International Conference on Phenomena in Ionized Gases (ICPIG), Bucharest, 1969*, p. 140.

18. A. A. Emel'yanov, Prib. Tekh. Éksp., No. 5, 68 (1997).

19. I. D. Chalmers and B. D. Phukan, Vacuum **32**, 145 (1982).

20. A. A. Emel'yanov, Prib. Tekh. Éksp., No. 6, 90 (1998).

Translated by M. Astrov

Theory of Surface Dynamic Variation during the High-Temperature Field Evaporation

V. N. Shrednik*, D. V. Glazanov**, and E. L. Kontorovich*

* Ioffe Physicotechnical Institute, Russian Academy of Sciences,
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
** Bonch-Bruevich State University of Telecommunications, ul. Moĭka 61, St. Petersburg, 191186 Russia
e-mail: glazanov@itcwin.com
Received February 13, 2003

Abstract—We consider the heated surface of a metallic tip to which a strong electric field is applied. At temperatures activating surface self-diffusion, crystalline outgrowths and microprotrusions arise on the surface. The latter generate ion fluxes, i.e., act as sources of high-temperature field evaporation, when a positive potential is applied to the emitter. The existence conditions for the microprotrusions on the emitter surface are discussed. It is shown that their stability is provided by the balance between three atomic fluxes: diffusion from the top of the tip, diffusion toward the top, and field evaporation from the top. Different ways of providing such a balance are discussed. In a desorption-type field ion microscope, the microprotrusions and evaporating ions are visualized as bright spots. These spots execute random motion and, at the same time, exhibit ordered cyclic displacements: the microprotrusions first form dotted rings along the developed faces of the crystalline emitter, and then these rings quickly collapse toward the center of the face. A quantitative theory of these cyclic processes is developed for the first time. It explains why the rings "calm down" before collapse and why subsequent collapse develops in an avalanchelike manner. The electric field distribution over the surface in the presence of an outgrowth is calculated, and diffusion fluxes at different stages of its growth and dissolution are analyzed. The calculation shows that the outgrowth heights are relatively small and their slopes are rather smooth. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The simultaneous action of field and temperature on the metallic tip surface results in a variety of phenomena, which modify the tip shape [1]. Such undesired thermal–field shaping should always be taken into account, since this process often causes the avalanche development of instabilities and adversely affects the emitter parameters. The detailed consideration of various instabilities on the surface of electron and ion emitters may form the basis for a theory and technology of reliable field sources of ions and electrons.

Among such surface tip-shaping phenomena is high-temperature field evaporation [2], which may be viewed as the process of ion evaporation accompanied by intense surface diffusion and the growth of microprotrusions and more complex features, such as outgrowths, steps, etc. It is remarkable that high-temperature field evaporation may be used as the principle of operation of point ion sources with a desired geometry and hence with emission properties that are necessary for the fabrication of micro- and nanoelectronic devices [3, 4].

Depending on the electric field and temperature, a wide variety of surface patterns may be observed: from a single microprotrusion on the surface of a regularly faceted crystal to many microprotrusions randomly arranged on the irregular "pitted" surface [1, 5]. The shape of the tip is very difficult to stabilize. As a rule, a number of dynamic processes related to crystal growth and dissolution are observed on the surface, which show up as the appearance and disappearance of spots in field-emission images.

Thermal–field processes and the occurrence of different surface features were studied experimentally for several refractory metals and alloys in a wide range of fields and temperatures. A desorption field-emission microscope was used as the main tool to observe high-temperature field evaporation [2, 5–7].

Nevertheless, several fundamental issues still remain unsolved. Namely, which process governs the number of microprotrusions on the top of the tip? Are these protrusions dynamically stable or basically equilibrium (or near-equilibrium)? What are the rules the motion of the microprotrusions on the surface and, accordingly, of emission spots in the images obeys? These issues are covered in this study.

TYPE AND NUMBER OF THERMAL–FIELD MICROPROTRUSIONS

Microprotrusions are relatively stable on most of the tip surface. They "twinkle" (fluctuate) but are observed over a long period of time. In general, this seems to be surprising, since so-called steady-state microprotru-
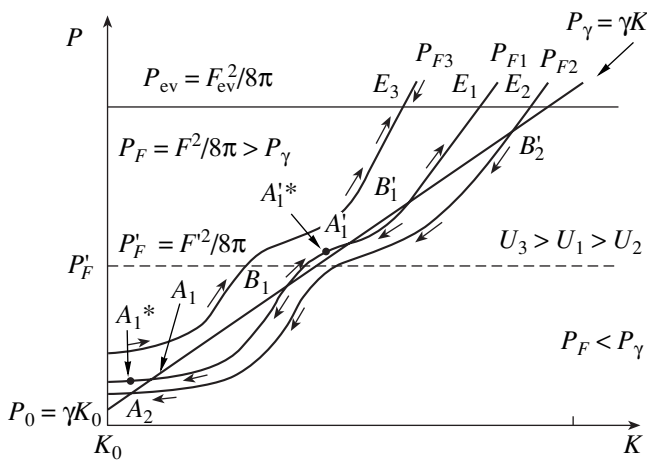
Fig. 1. Negative field pressure P_F at the emitter top vs. the surface curvature K. The curves are drawn for three values of the voltage U, the straight line depicts the surface tension force pressure $P_\gamma(K) = \gamma K$, and the line P_{ev} reflects field evaporation that balances field-induced growth.

sions (i.e., those provided by the balance between field evaporation and field supply) must not apparently be very stable.

Figure 1 shows a segment of the *P–K* diagram, where P is the pressure and K is the tip curvature, which was comprehensively analyzed in [5]. The diagonal straight line of surface tension force pressure $P_\gamma(K) = \gamma K$ (γ is the surface tension coefficient) divides the diagram into two parts. In the upper left part, the electric field pressure $P_F = F^2/8\pi$ exceeds P_γ and the field growth (indicated by the arrows in Fig. 1) causes the tip to sharpen. In the lower right part, $P_F < P_\gamma$ and the natural trend is blunting.

Figure 1 presents the $P_F(K)$ curves for three values of the voltage U. The line $P_{F1}(K)$ for voltage U_1 crosses the diagonal boundary line $P_\gamma(K)$ four times at points of equilibrium, only two of which (points of type A, see below) correspond to stable equilibrium. These are A_1, which corresponds to the shape of the faceted reconstructed tip, and A_1', which corresponds to a microprotrusion on a vertex of a polyhedron. The curve $P_{F2}(K)$ for a voltage U_2 lower than U_1 has only two points of intersection with the line $P_\gamma(K)$ and one point of stable equilibrium, A_2, in the case of the reconstructed shape. The curve $P_{F3}(K)$ ($U_3 > U_1$) does not cross the demarcation line at all. The only stable state in this case is associated with the point E_3 on the line P_{ev}, where the diffusion and field evaporation fluxes compensate for each other. The undulatory of the curves $P_F(K)$ is the superposition of two parabolas (for details, see [5]). All the curves in Fig. 1 correspond to a certain constant temperature T. Note that the surface tension coefficient γ and, accordingly, the values of P_γ and P_{ev} are temperature dependent.

As follows from Fig. 1, a high field F (and, hence, high P_F) causes field evaporation, which compensates for the atomic flux toward the tip along the line P_{ev} at stationary E points. The fluctuations of the field F (which may be caused by the natural instability of the voltage U) displace (in terms of Fig. 1) the microprotrusion to the nearest curve P_F. In this case, random increases in F somewhat blunt the tip; nevertheless, the microprotrusion remains stationary. However, random fluctuation decreases in F sharpen the microprotrusion, displacing the point E to the right. A series of such decreases in F may lower the point in the *P–K* diagram below the line P_γ. Then the tip shape will rapidly and irreversibly change: the tip will become blunted, and a sharp microprotrusion will turn into the basic reconstructed form. Such an upset is described by motion along the curve $P_2(K)$ approximately from the point B_2' to the point A_2 of stable equilibrium. For this reason, microprotrusions corresponding to the points E cannot be long-lived. At the same time, they are expected to provide the highest ion current densities during high-temperature field evaporation.

Numerous relatively stable microprotrusions observed under a desorption field microscope are probably not stationary but near-equilibrium. Equilibrium microprotrusions (corresponding to the points A') may not evaporate the material. However, depending on the temperature and material of the tip, the field F may be large enough to trigger field evaporation at the point A' (or even at the point A). In this situation, the balance between atom fluxes toward and from the top of the tip (the latter is due to surface diffusion in the field) certainly breaks. The field evaporation flux adds up to the backward flux caused by P_γ. Accordingly, the tip gets blunted. In the meantime, the difference $(P_F - P_\gamma)$, which governs the atom flux to the top, will increase and the flux of field evaporation will decrease with decreasing F. Such a tendency will persist until the atom flux to the top due to the field becomes equal to the sum of the backward fluxes blunting the protrusion: the flux due to P_γ and that associated with field evaporation. A new stability point will be situated in the curve P_F somewhat to the left from the point A' (or A) and will correspond to a stable, though weak, flux of field evaporation with the emitter shape retained (self-sustained). It is this situation that we characterize as corresponding to a near-equilibrium microprotrusion (point A'^*) or a near-equilibrium top of the tip (point A^*). In this case, the fluctuations of the voltage (field F) do not result in an avalanchelike decay of the microprotrusion. The vortex curvature of the microprotrusion (as for point A) follows the variation of F: grows with increasing F and decreases when the field decreases. When the emitter shape fluctuates (i.e., K varies), the surface geometry returns to point A'^* or A^*, which is typical of points A and A' without evaporation.

It appears that there is a threshold field F' of formation of thermal–field protrusions that depends on the material and temperature of the emitter [8]. This threshold also depends on the applied voltage and the tip shape and can be achieved at one point or in a small number of points on the surface. In the latter case, we will observe a small number of microprotrusions. If F' is reached at many sites, a great number of microprotrusions and outgrowths will form and the surface will be highly irregular. Obviously, thermal–field microprotrusions will initially arise on the surface of the reconstructed tip at edges and vertices formed by the intersection of crystal faces. If surface areas with the highest K evaporate ions, the microprotrusions will be either near-equilibrium (point A'^*) or stationary (point E). Their character will depend on the position of the curve $P_F(K)$ relative to the line P_v and on the position of point P_F' corresponding to F' on curve $P_F(K)$. When moving along curve $P_F(K)$ to the right from point (P_F', K'), the representation point corresponding to the growing emitter arrives at either point A'^* (e.g., on curve P_{F1}) or (if point A is absent) directly at point E (as in curve P_{F3}).

MACROOUTGROWTHS AND THE DYNAMICS OF THEIR GROWTH AND DISSOLUTION

If the temperature and field are high enough, macrooutgrowths grow on large close-packed faces [1]. Their transverse sizes are comparable to the size of the face and depend on the curvature of the tip base, reaching 1000 Å. These macrooutgrowths look like truncated cones or pyramids. Their height has not yet been reliably established. The estimation of this value is one of the goals of this study. As was noted in [6, 9], the growth of macrooutgrowths corresponds to the slow stage of formation of a bright ring from spots due to microprotrusions, which borders the top of the tip.

In [6, 7], the fast collapse of the rings upon high-temperature field evaporation was observed. At this stage, outgrowths narrow in an avalanchelike manner. Under a desorption field-emission microscope, the rings were observed for evaporating ions of Ir, Pt, and others [6, 7, 10]. The (dotted) rings consist of individual spots corresponding to thermal–field microprotrusions (Fig. 2) and fringe close-packed faces of the tip base. For Ir and Pt, these are {111} and {100} faces.

When localized at the edge of a close-packed face fluctuating (arising and disappearing) microprotrusions set conditions for the layer growth of this face, i.e., serve as new-layer nuclei [11]. Apparently, a limited number of microprotrusions around a close-packed facet cannot provide its rapid growth. When the number of microprotrusions becomes sufficiently large (in the image, they entirely fringe the face, forming a dotted ring), oriented stable layer growth takes place. As long as the height of a macrooutgrowth is one or two monolayers, the field near the exterior angle at the base of the
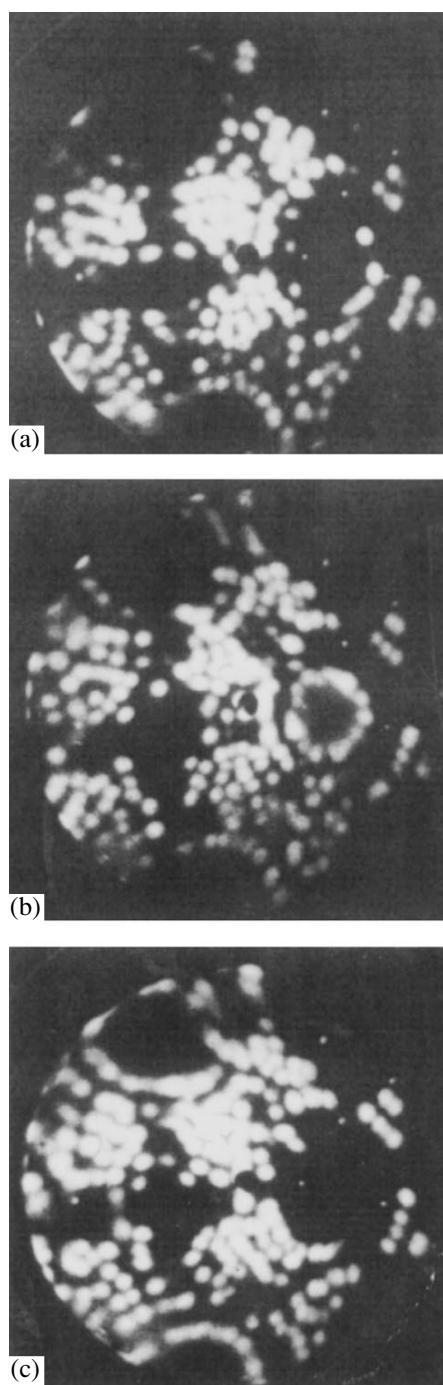
Fig. 2. Different stages of the high-temperature field evaporation of iridium single crystal with many microprotrusions [6]. $U = 9.9$ kV, $T = 1340$ K. Microprotrusions corresponding to individual spots are imaged in a field desorption microscope by means of Ir ions fluxes from their surface. During the observation, the microprotrusions (spots) twinkle and move. (a) Spots start to form a ring around {111} faces (on the right and at the bottom), (b) the ring on a {111} face on the right has been formed and is ready to collapse, (c) the ring disappeared (the stage after the collapse). Rings formed on the top and bottom {111} faces are clearly seen.

Fig. 3. Field distribution over the outgrowth surface.

cone (where the surface curvature is negative), while starting to drop, remains high enough to pull atoms in the growth region. As the growth continues, the field at the exterior angle continues to drop and eventually reaches the level where the atom flux 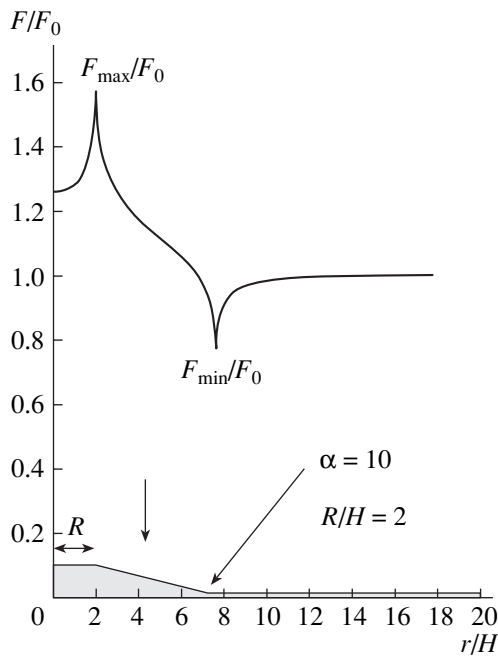toward the top is limited because of the reduction of the chemical potential gradient [12]. It appears that the ring "calms down" for a time from tens of seconds to several minutes just in this situation [6]. At this stage, the field at the top of the cone is enhanced and favors intense field evaporation. Under these conditions, the material flux for further growth is provided mainly by atoms from the lateral surface of the cone. As a result, the transverse size of the cone and the diameter of its base shrink. The field around the top of the tip increases, causing the intensification of the evaporation, a further decrease in the cone diameter, and, accordingly, an increase in the field at the upper edge. Eventually, the avalanche process of macrooutgrowth evaporation develops, showing up as ring collapse in the image.

This physical picture of collapse of rings consisting of microprotrusions was presented in [9]. For clarity, the cones of microprotrusions in [9] were depicted high with steep slopes. Note that no quantitative analysis of the model proposed was suggested.

CALCULATION OF THE FIELD DISTRIBUTION OVER THE TIP SURFACE

In order to refine the above qualitative picture, we numerically calculated the electric field distribution over the tip surface with a macrooutgrowth. In the calculations, the outgrowth was approximated by a trun-

cated cone. Since the radius of surface curvature of the tip bases in experiments [6, 7] was about 1 μm, we assumed this surface to be planar. The field was studied as a function of two parameters: the base angle α of the macrooutgrowth and the ratio R/H, where R is the radius of the upper base of the outgrowth and H is its height.

The Laplace equation for the potential in the interelectrode space was solved by the iterative finite-difference method on a nonuniform mesh using successive overrelaxation. To calculate the electric field distribution over the macrooutgrowth surface and the remaining tip surface, we approximated the potential values by a two-dimensional cubic spline, since the direct application of the finite-difference method to calculating the potential gradient does not provide the smoothness required. It was also assumed that angles on the vertical-plane section of the cone are not rounded. This means that the actual radii of curvature of the edges are smaller than the characteristic size of the mesh. This is the case if the edges and vertices of the macrooutgrowth are atomically (or nearly atomically) sharp.

Obviously, there are two extreme lines: the edge of the upper surface of the macrooutgrowth, where the field is maximal (F_{max}), and the boundary of its base, where the field is minimum (F_{min}). Figure 3 shows the distribution of the ratio F/F_0 over the surface, where F_0 is the field away from the macrooutgrowth. The calculations show that the field contrast $\delta = F_{max}/F_{min}$ depends strongly on the base angle α of the cone (Fig. 4). The ratio δ is practically independent of the ratio of the height H to the transverse size of the cone, e.g., to the radius of the upper base R. This statement is valid for low outgrowths ($R \gg H$). However, for not too small angles α ($\alpha > 5°$), δ starts to increase rapidly as R decreases to values on the order of H or below (Fig. 5). Such behavior strongly supports the avalanche mechanism of ring collapse discovered in [6].

However, reasons why the growth of the macroprotrusion is suppressed and even ceases before the ring collapses deserve special consideration. According to [12], the flux of atoms diffusing over the surface is proportional to the chemical potential gradient $\nabla\mu = \nabla(\gamma K - F^2/8\pi)$ and is directed opposite to the gradient, i.e., toward the region where the field F increases and the surface curvature K decreases. When approaching the bottom of the outgrowth, the field decreases (Fig. 3) and the related component $(\nabla\mu)_1$ of the gradient μ retards the flux toward the outgrowth. However, it should be taken into account that the area around the outgrowth base (which is assumed to be flat in our calculations) is actually a part of the tip base, which has a noticeable positive curvature. Near the base, the positive curvature changes to negative curvature (initially the upper edge of the tip has zero curvature). Such a geometry and the related electric field distribution defines a certain initial value of $(\nabla\mu)_0$. When $(\nabla\mu)_1$ at the point of minimum field equals $(\nabla\mu)_0$, the atomic
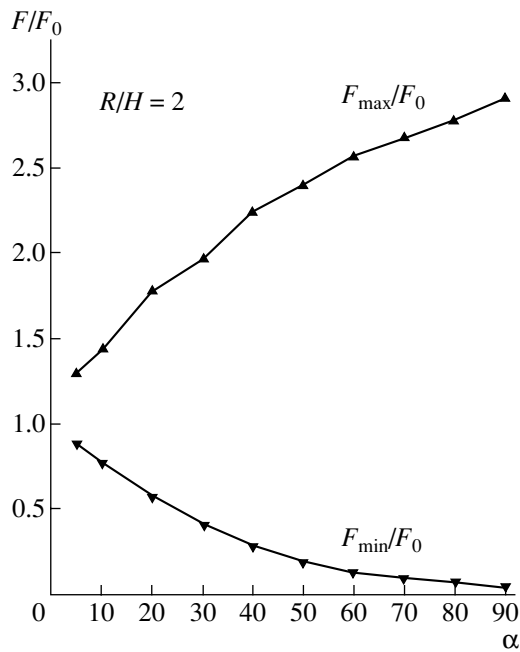
Fig. 4. Field intensity at the upper edge and near the base of the outgrowth as a function of the base angle. F_0 is the field on the surface away from the macrooutgrowth.
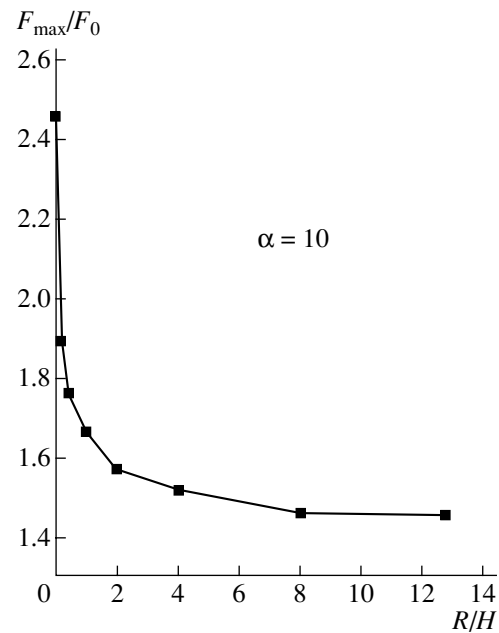

Fig. 5. Field intensity at the upper edge of the outgrowth as a function of the base-to-height ratio of the outgrowth. F_0 is the field on the surface away from the outgrowth.

flux from the tip surface toward the lateral surface of the outgrowth stops. However, since the field F_{max} and high temperature stimulate intense field evaporation from the upper edge of the outgrowth, the process will involve the material of the lateral surface. As a result, the transverse size of the outgrowth decreases and the ring collapses.

The results of the field calculations suggest that the field contrast δ exceeds two even if α and ratio H/R are not very high. At the same time, it appears that $\delta \approx 2$ would apparently suffice for the experimental observation of such effects as the suppression of the diffusion flux from the tip base (showing up as the slow growth of the protrusion and ring stabilization) or a drastic enhancement (initiation) of field evaporation, since the rates of diffusion and evaporation depend exponentially on the field F. The dependences of F_{max} and F_{min} on the outgrowth base angle α are presented in Fig. 4.

From this figure, it follows that the lateral slopes of forming macrooutgrowths must be smooth (α is no more than $10°–15°$, Fig. 4) and their heights must be small (several atomic layers), since F_{max} increases with height H (Fig. 5), enhancing material evaporation. This refers to outgrowths for which ring collapse is observed.

The experimental data [6, 7] support this conclusion: in field-emission images, microprotrusions in the ring before its collapse are similar to many others outside the ring. If, for example, $\alpha \approx 60°$, the values of δ are so large that the growth of protrusions may be suppressed from the very beginning. Quantitatively, this

conclusion disagrees with the scheme of the process presented in [9] but supports the underlying mechanism.

When observed in the field-ion regime [1], the outgrowths do have a smooth shape when initially many microprotrusions are present on the surface. If the outgrowths were high, the microprotrusions (grouped into rings) would be observed only along the circumference of the outgrowth. However, during the avalanche collapse of the rings, the emission pattern obtained with Ir and Pt ions [6, 7] was considerably brighter than expected. This points to a noticeably enhanced evaporation rate and a high ion-current density, which suggests the evaporation of stationary microprotrusions (point E in Fig. 1). The ratio of the fields in the upper and bottom angles (up to 100) for a high protrusion would be considerably higher than that at which the development of macrooutgrowths switches to its intense evaporation, i.e., to the stage of ring collapse.

CONCLUSIONS

(1) The number of thermal–field microprotrusions depends mainly on the surface condition (irregularity). The surface condition is specified by the history of thermal–field treatment, instantaneous values of temperature and field, and the emitter material.

(2) Thermal–field microprotrusions, which are usually observed *in situ* in the desorption regime, are most probably near-equilibrium, rather than stationary, features.

(3) The collapse kinetics of rings made up of micro-protrusions growing around the upper face of the macrooutgrowth consists in suppressing the atomic flux as the outgrowth develops and, accordingly, reducing the field near its base.

(4) The calculations show that the collapse of the rings during high-temperature field evaporation correlates with the dissolution of relatively low outgrowths (several atomic layers) with smooth slopes ($10°$–$15°$).

ACKNOWLEDGMENTS

REFERENCES

1. V. N. Shrednik, in *Crystal Growth* (Nauka, Moscow, 1980), Vol. 13, pp. 68–79.

2. Yu. A. Vlasov, V. G. Pavlov, and V. N. Shrednik, Pis'ma Zh. Tekh. Fiz. **12**, 548 (1986) [Sov. Tech. Phys. Lett. **12**, 224 (1986)].

3. V. N. Shrednik, in *Review of the 7th International Conference on Vacuum Microelectronics "Le vide les Couches Minces," Grenoble, 1994*, Suppl. 271, pp. 420–421.

4. V. N. Shrednik, Mikroélektronika **26** (2), 97 (1997).

5. Yu. A. Vlasov, O. L. Golubev, and V. N. Shrednik, Izv. Akad. Nauk SSSR, Ser. Fiz. **52**, 1538 (1988).

6. V. G. Butenko, O. L. Golubev, E. L. Kontorovich, and V. N. Shrednik, Pis'ma Zh. Tekh. Fiz. **18**, 86 (1992) [Sov. Tech. Phys. Lett. **18**, 275 (1992)].

7. O. L. Golubev, E. L. Kontorovich, and V. N. Shrednik, Zh. Tekh. Fiz. **66** (3), 97 (1996) [Tech. Phys. **41**, 284 (1996)].

8. V. G. Pavlov, A. A. Rabinovich, and V. N. Shrednik, Zh. Tekh. Fiz. **47**, 405 (1977) [Sov. Phys. Tech. Phys. **22**, 239 (1977)].

9. O. L. Golubev and V. N. Shrednik, in *Proceedings of the 16th International Symposium on Discharges and Electrical Insulation in Vacuum, St. Petersburg, 1994*, pp. 473–476.

10. M. V. Loginov and V. N. Shrednik, Zh. Tekh. Fiz. **67** (8), 102 (1997) [Tech. Phys. **42**, 1075 (1997)].

11. Yu. A. Vlasov, O. L. Golubev, and V. N. Shrednik, in *Crystal Growth* (Nauka, Moscow, 1991), Vol. 19, pp. 5–21.

12. C. Herring, *Structure and Properties of Solid Surfaces* (Chicago Univ. Press, Chicago, 1953), pp. 5–81.

Translated by M. Lebedev

BRIEF
COMMUNICATIONS

On the Current through a Virtual Cathode

A. E. Dubinov and I. A. Efimova

Institute of Physics and Technology, Sarov, Nizhegorodsk oblast, 607188 Russia
e-mail: dubinov@ntc.vniief.ru
Received October 21, 2002

Abstract—Two existing opinions about the current passing through a virtual cathode in a closed equipotential cavity are considered: (i) the current is always smaller than half the limiting current (as follows from the steady-state consideration) and (ii) the current oscillates in time about the mean value that is exactly equal to the limiting current (as follows from the non-steady-state self-consistent simulation by the coarse particle method). The history of this issue is described. The invalidity of the former opinion is shown. © *2003 MAIK "Nauka/Interperiodica".*

Electron beams with a virtual cathode (VC) are finding increasing application in many fields of technology, such as the generation of bremsstrahlung and microwave radiation, the collective acceleration of electrons and ions, high-current switches, etc. A review of the state of the art in the development of VC devices can be found in [1, 2].

An important issue concerned with VC device operation is the value of the current I_{vc} passing through the VC. The problem is stated as follows. Let a monoenergetic magnetized rectilinear electron beam with a time-constant injection current I_{inj} enter into a closed equipotential cavity of arbitrary shape and let the injection current exceed the limiting current I_{lim} (so that a VC is formed). The goal is to find the value of the current passing through the VC.

There are two opinions about the value of this current. We will try to elucidate which of them is the correct one.

The former opinion follows from the analysis of the steady states of the electron beam in the cavity. It shows that, when intense electron beams are injected into equipotential cavities of different configurations, the nonlinearity due to the field produced by the self-space charge of the beam causes an intriguing effect: the beam may take two states with its parameters remaining the same (state hysteresis). One of them (steady state) corresponds to the case where the entire beam crosses the cavity; the other, to the case where some of the electrons reflect from a certain section. Such hysteretic behavior has been known since 1923 [3].

Under such conditions, many of the parameters that characterize the beam's state exhibit a two-valued dependence on the beam current [4–7] (in a certain range of the current). This phenomenon may be used for generating high-power nanosecond current pulses [8, 9], as well as for the collective acceleration of positive ions [10, 11]. In addition, the hysteresis phenomenon may be applied to advantage for generating high-power electromagnetic oscillations of relaxation type [12, 13].

The stationary analysis of the problem for a cavity in the form of a plane gap in the nonrelativistic case shows that, when the injection current I_{inj} varies, the current I_{vc} having passed through the virtual cathode also exhibits hysteresis (Fig. 1). Its descending branch *2* corresponding to the VC regime is described by the dependence

$$\frac{I_{vc}}{I_{lim}} = \frac{I_{inj}}{I_{lim}}$$
$$\times \left(1 - \sqrt{1 - \frac{1}{2(I_{inj}/I_{lim})} - \frac{1 + \sqrt{1 + 8(I_{inj}/I_{lim})}}{8(I_{inj}/I_{lim})^2}} \right). \quad (1)$$

Such a notion of the behavior of I_{vc} is generally accepted and is common in the original scientific and technical papers [6, 14, 15], monographs [16–19], textbooks [4, 20, 21], etc.
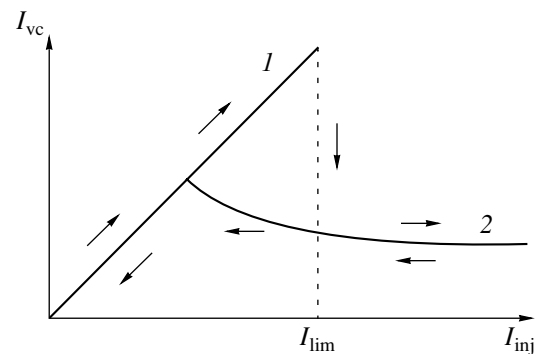

Fig. 1. Hysteretic dependence of the passing current I_{vc} on the injection current I_{inj}: (*1*) passage of all the beam's electrons and (*2*) the VC regime.

The plot in Fig. 1 is treated as follows. As the injected current I_{inj} increases slowly, so does the current I_{vc} (since $I_{vc} = I_{inj}$) according to upper branch *1*. When I_{inj} reaches the limiting value I_{lim}, the current drops stepwise to a value smaller than $I_{lim}/2$ following descending branch *2*. From this point on, a VC forms in the beam and, as the current I_{inj} rises further, the passing current tends to $I_{lim}/8$. When the current I_{inj} decreases, I_{vc} increases along branch *2*. Thus, it is believed that

$$I_{vc} \in (I_{lim}/8,\ I_{lim}/2) < I_{lim}. \qquad (2)$$

Strict inequality (2) is central to the first opinion about the current passed through a VC. Note that, in terms of the steady-state consideration, the behavior of a relativistic electron beam is qualitatively similar; however, exact characteristic values cannot be found in this case. The behavior of electron beams in equipotential cavities of another shape has also been found to be nearly the same (for example, the behavior of a thin-walled tubular beam in a cylindrical tube under longitudinal coaxial injection [22–24]).

Yet numerical simulation by the coarse particle method yields another result: the current through a VC oscillates in time about I_{lim}, so that we can write

$$\langle I_{vc} \rangle = I_{lim} \qquad (3)$$

with a high accuracy (within 2%).

In the simulation, we used the 2.5-dimensional version of the particle-in-cell Karat code [25], which allows for the self-consistent solution of the Maxwell equation on a mesh and relativistic equations of motion of particles. The code was repeatedly employed to advantage in solving similar problems by researchers throughout the world. Figure 2 shows the simulated
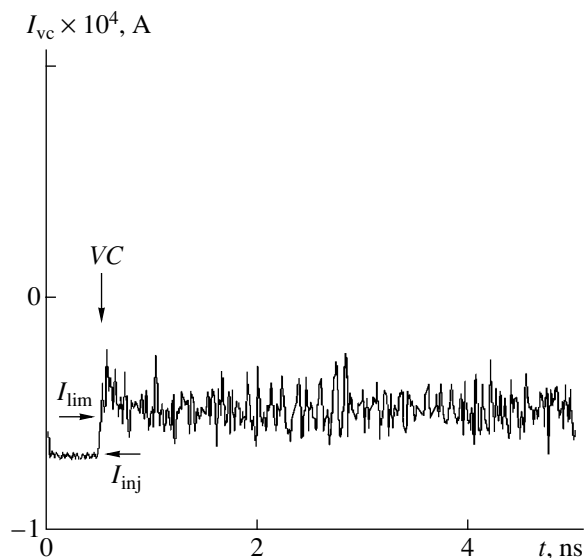

Fig. 2. Calculated waveform of the passing current I_{vc} [26]. Horizontal arrows indicate the levels of I_{inj} and I_{lim}; the vertical arrow points to the instant of VC occurrence.

waveform of the passing current [26], which illustrates (3). The early portion of the waveform corresponds to those values of I_{inj} when a VC has not formed yet, while the remaining part of the waveform, after the VC has formed, shows intense oscillations of I_{vc} about the mean value equal to I_{lim}.

It turned out that the same result was also obtained in works employing alternative codes based on the coarse particle method. For example, it was reported [27] that I_{vc} and I_{lim} for a beam with a VC differ by no more than 12% (the originally developed code was used). In [28], with the well-known particle-in-cell Magic code [29], it was demonstrated that equality (3) holds with an accuracy of as high as 1%. The authors of [29] also discovered strong oscillations of I_{vc}. In [30] (original code), the mean passing current exceeded the critical value by only 0.75%. In [31], it was found that $\langle I_{vc} \rangle = 0.96 I_{lim}$ or $0.975 I_{lim}$ depending on the problem parameters selected (also the original code). Other works supporting the validity of (3) are also known.

In view of the high confidence of equality (3), we used it to develop a simple model of induced radiation in a vircator [32] and also for further testing the Karat code [26].

Equality (3) forms the basis for the second opinion about the value of the current passing through a VC.

Thus, we have presented two mutually exclusive opinions concerning the value and behavior of I_{vc}, which are based on (2) and (3). Both have many advocates. The point of the value of the passing current is of crucial importance for microwave devices, as well as for the collective acceleration of electrons and ions, and should be resolved as soon as possible. It has become the subject of hot discussion [33, 34]. The authors of the present work hold to the second opinion.

To prove the invalidity of inequality (2), we will turn to [35], where this inequality was derived for the first time from the nonstationary Poisson equation for electrical potential. At the point where a stationary VC has been formed, the electrical potential has a minimum, the electron velocity equals zero, and the beam is divided into transmitted and reflected parts. In other words, the gap is conventionally partitioned into two diodes: a double-flow diode before the VC and a Chaild–Langmuir diode after the VC. The position of the VC is thus defined by the balance of the currents in these diodes, which is reflected in (1) and (2). In all the papers listed above, the derivation drawn in [35] is repeated.

However, the major difficulty associated with the derivation of (2) was noted only in the original work [35]. In subsequent papers, it was disregarded. The question arises as to why and how some of the electrons reflect from the potential barrier, while others pass through it. Recall that the beam injected is monochromatic. Lukoshkov [35] notes that "the assumption that there is a minor, e.g., thermal, electron velocity spread

suffices to explain the mechanism of separating the electrons into transmitting and reflecting ones." However, this stipulation comes into conflict with the statement of the problem. At the same time, it is clear that the ratio of the transmitted and reflected electrons must depend primarily on the electron velocity distribution form. Thus, an uncertainty in the transmitted-to-reflected electron ratio (more strictly, the impossibility of its correct explanation) is the fundamental disadvantage of the steady-state model.

Numerical nonstationary simulation by the coarse particle method, which cannot be performed at the time [35] appeared, immediately revealed that a VC is a basically nonstationary object: its potential and spatial position oscillate in time and the oscillations are not small [36]. It turned out that a VC acts as a valve, transmitting all electrons of a monoenergetic beam in one phases of the oscillations and reflecting them back in others. Such a beam splitting mechanism has been called phase separation [37]. Phase separation is responsible for radiation instability on a vircator [32]. Advanced facilities of computational experiment allow one to dynamically visualize the process of phase separation.

To conclude, we note that the various types of VCs were first classified in review [1]. According to this classification, the process considered by us has to do with the dynamics of an oscillating VC, which has a threshold in current and may arise in a monochromatic beam. Lukoshkov [35] tried to replace certain signs of an oscillating VC by those of a steady-state thermal VC, which has no current threshold, cannot exist in a monoenergetic beam, and demonstrates the radically different kinetics.

Thus, relationships (2) is invalid and the conclusions drawn in [33] are in error. When designing devices with a VC, one must be guided by relationship (3), which has been proved many times in self-consistent calculations and in experiments [38].

ACKNOWLEDGMENTS

REFERENCES

1. A. E. Dubinov and V. D. Selemir, Radiotekh. Élektron. (Moscow) **47**, 645 (2002).

2. A. E. Dubinov, I. Yu. Kornilova, and V. D. Selemir, Usp. Fiz. Nauk **172** (11), (2002) [Phys. Usp. (2002)] (in press).

3. V. Bursian and V. Pavlov, Journal of the Russian Physico-Chemical Society **55** (1–3), 71 (1923).

4. S. D. Gvozdover, *Theory of Microwave Electronic Devices* (GITTL, Moscow, 1956).

5. A. V. Pashchenko, B. N. Rutkevich, V. D. Fedorchenko, and Yu. P. Mazalov, Zh. Tekh. Fiz. **53**, 75 (1983) [Sov. Phys. Tech. Phys. **28**, 42 (1983)].

6. C. E. Fay, A. L. Samuel, and W. Shockley, Bell Syst. Tech. J. **17** (1), 49 (1938).

7. A. E. Dubinov, Vopr. At. Nauki Tekh., Ser.: Teor. Prikl. Fiz., No. 2, 3 (2001).

8. A. V. Pashchenko and B. N. Rutkevich, USSR Inventor's Certificate No. 646783; Byull. Izobret., No. 30 (1979).

9. A. E. Dubinov, N. N. Makarova, and V. D. Selemir, RF Patent No. 2155411; Byull. Izobret., No. 20 (2001).

10. A. G. Lymar', V. V. Belikov, A. V. Zvyagintsev, *et al.*, Vopr. At. Nauki Tekh., Ser.: Yad. Fiz. Issled., No. 5, 71 (1989).

11. A. E. Dubinov, RF Patent No. 2171017; Byull. Izobret., No. 24 (2000).

12. M. N. Devyatko, Elektron. Tekh., Ser. 1: Elektron. SVCh, No. 5, 75 (1979).

13. Yu. K. Alekseev and A. I. Kostienko, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **21** (2), 95 (1980).

14. A. A. Kolomenskiĭ, in *Proceedings of the Meeting on Problems of the Collective Method of Acceleration, Dubna, 1982* (Ob"edin. Inst. Yad. Issled., Dubna, 1982), pp. 71–80.

15. C. K. Birdsall, in *Proceedings of the Symposium on Plasma Double Layers, RISO National Laboratory, Roskilde, 1982*, Ed. by P. Michelsen and J. J. Rasmussen, p. 274.

16. L. N. Dobretsov and M. V. Gomoyunova, *Emission Electronics* (Nauka, Moscow, 1966).

17. R. B. Miller, *Introduction to the Physics of Intense Charged Particle Beams* (Plenum, New York, 1982; Mir, Moscow, 1984).

18. I. N. Meshkov, *Transport of Charged Particle Beams* (Nauka, Novosibirsk, 1991).

19. C. K. Birdsall and W. B. Bridges, *Electron Dynamics of Diode Regions* (Academic, New York, 1966).

20. A. I. Lebedev, *Physical Processes in High-Current Diodes* (Mosk. Inzh.-Fiz. Inst., Moscow, 1984).

21. Yu. K. Alekseev and A. P. Sukhorukov, *Introduction into the Theory of Catastrophes* (Mosk. Gos. Univ., Moscow, 2000).

22. S. P. Bugaev, N. I. Zaĭtsev, A. I. Kim, V. I. Koshelev, A. I. Fedotov, and M. I. Fuks, *Relativistic High-Frequency Electronics* (Inst. Prikl. Fiz. Akad. Nauk SSSR, Gor'kii, 1981), pp. 36–61.

23. M. I. Fuchs and N. F. Kovalyov, in *Proceedings of the 10th Conference on Controlled Fusion and Plasma Physics, Moscow, 1981*, Vol. 1, p. M-4.

24. N. F. Kovalev and M. I. Fuks, Zh. Tekh. Fiz. **52**, 2002 (1982) [Sov. Phys. Tech. Phys. **27**, 1230 (1982)].

25. V. P. Tarakanov, *User's Manual for Code Karat* (Berkley, Springfield, 1992).

26. A. E. Dubinov and I. A. Efimova, Zh. Tekh. Fiz. **71** (6), 80 (2001) [Tech. Phys. **46**, 723 (2001)].

27. V. P. Grigor'ev, V. P. Il'in, T. V. Koval', N. I. Sablin, and M. V. Urev, *Preprint of the Computer Center, Siberian Division, Academy of Sciences of USSR*, Novosibirsk, 1990, no. 889.

28. Y. Seo, E. H. Choi, C. G. Byun, and M. C. Choi, in *Proceedings of the 13th International Conference on High-Power Particle Beams (BEAMS 2000), Nagaoka, 2000*, pp. 742–745.

29. B. Goplen, L. Ludking, and D. Smithe, *Magic User's Manual* (Mission Res., Newington, 1996), Vol. MRC/WDC-R-380.

30. V. A. Vshivkov, Chislennye Metody Mekh. Sploshnoĭ Sredy **10** (2), 40 (1979).

31. D. S. Filippychev and V. F. Shafer, Fiz. Plazmy **15**, 182 (1989).

32. A. E. Dubinov, V. D. Selemir, and A. V. Sudovtsev, Elektron. Tekh., Ser. 1: Elektron. SVCh, No. 1 (469), 7 (1997).

33. N. F. Kovalev, Zh. Tekh. Fiz. **72**, 113 (2002) [Tech. Phys. **47**, 906 (2002)].

34. A. E. Dubinov and I. A. Efimova, Appl. Phys. Lett. **81**, 1530 (2002).

35. V. S. Lukoshkov, Zh. Tekh. Fiz. **6**, 624 (1936).

36. W. B. Bridges and C. K. Birdsall, J. Appl. Phys. **34**, 2611 (1963).

37. V. D. Selemir, A. B. Alekhin, V. E. Vatrunin, *et al.*, Fiz. Plazmy **20**, 689 (1994).

38. G. P. Mkheidze, V. I. Pulin, M. D. Raĭzer, and L. É. Tsopp, Zh. Éksp. Teor. Fiz. **63**, 104 (1972) [Sov. Phys. JETP **36**, 54 (1972)].

Translated by V. Isaakyan

<div align="center">

BRIEF COMMUNICATIONS

</div>

Interpretation of Velocity Measurements as Indirect in the Trajectory Finding Problem

A. S. Devyatisilny

Institute of Automation and Control, Far East Division, Russian Academy of Sciences, Vladivostok, 690041 Russia
e-mail: devyatis@iacp.dvo.ru
Received December 17, 2002; in final form, March 5, 2003

Abstract—It is shown that the necessary condition for the correct statement of the inverse trajectory problem in measuring the linear velocity vector of a material point in the gravitational field is the nondegeneracy of the matrix, i.e., the Hessian of the field potential function. © *2003 MAIK "Nauka/Interperiodica".*

It is known that knowledge of the velocity vector is of key importance for finding a trajectory.

Let direct measurements (obeying the basic physical principles [1]) of the absolute linear velocity (V) vector of a unit-mass material point be feasible in an orthogonal inertial frame of reference $0\xi = 0\xi_1\xi_2\xi_3$ with the origin at the point 0. Then, one can find the radius vector R of the point in this frame of reference at any time instant $t \geq t_0$ by directly integrating the equation (differential relation)

$$V = \dot{R} \tag{1}$$

provided that the initial position of the point, $R(t_0) = R_0$, is known. It is, however, obvious that the entropy of such a solution will increase with time because of inevitable perturbations (errors in determining the initial vector R_0, time, and velocity).

Let us state the problem of finding the vector R alternatively. We will consider the measurements of V as indirect measurements carrying information on R. In essence, this means that we consider the problem inverse to (1), bearing in mind the same purpose: to find R. Such a statement requires additional conditions to be imposed on V that do not contradict the established physical concepts of the nature od mechanical motion.

In terms of classical mechanics [2], these conditions are generally well known. In the form of cause–effect relation, they are given by Newton's second law

$$\dot{V} = a, \tag{2}$$

where a is the specific force acting on the point. We assume that a is a vector function continuous in R; that is, $a = a(R)$.

Let us write (1) and (2) jointly in the small:

$$\delta V = \delta \dot{R}, \quad \delta \dot{V} = a'\delta R, \tag{3}$$

where $a' = \partial a / \partial R$.

Set (3) is remarkable in that it is totally consistent with the approach used when the problem of inverse problem solvability is treated as a problem of observability in terms of the general notions of a system [3].

In fact, as follows from the above, the matrix $\begin{pmatrix} E & O \\ O & a' \end{pmatrix}$ for the vector $(\delta \dot{R}^T, \delta R^T)^T$ in set (3) (E is the unity matrix and T means the transposition of the vectors) is the observability matrix. Hence, the matrix a' must be nonsingular almost everywhere in the solution range.

The physical reality is such that the force $a(R)$ may be represented as

$$a(R) = \partial U / \partial R + f,$$

where U is the gravitational field potential (the function continuous in R) and f is the vector of specific nongravitational forces (hereafter, we put $f = 0$, bearing in mind that the vector f can be measured directly in a number of practical cases, for example, with a spatial newtonmeter [2] and thus can be included in a solution).

In view of the above, the matrix a' is identical to the Hessian of the gravitation potential function (that is, $a' = U''$). If the real function $U(R)$ is continuous in R, the Hessian is a real symmetric function.

Let the field be central; i.e., $U = \mu/|R|$, where μ is the gravitational parameter of the center. Such a model is common in describing the exterior field of terrestrial gravitation (because the central component prevails in the expansion of the terrestrial potential). Then, the matrix

$$U'' = -\frac{\mu}{|R|^3}\left(E - \frac{3RR^T}{|R|^2}\right)$$

is of full rank (with the singular numbers in the ratio of 2 : 1 : 1) and the problem considered is basically solvable.

Consider relative measurements. Let there be a Cartesian system $ox = ox_1x_2x_3$ rotating with a given angular

velocity $\omega = (\omega_1, \omega_2, \omega_3)^T$ and let us measure the projections of the relative linear velocity vector $v = \dot{r}$ of a material point onto the Cartesian axes (r is the radius vector of the material point in the projections onto the axes of the trihedron ox). Then, instead of set (3) in the small, we have the set

$$\delta v = \delta \dot{r},$$

$$\delta \dot{v} = 2q\delta \dot{r} + (\dot{q} + q^2 + U'')\delta r, \qquad (4)$$

where

$$U'' = \partial^2 U / \partial r^2; \quad q = \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix}.$$

As follows from (4), the statement of the problem is correct if the matrix $\dot{q} + q^2 + U''$ is nonsingular almost everywhere in the solution range.

Assuming again that the gravitational field is central and $f = 0$, we consider the special case of the rotation of the trihedron ox where $\omega = \text{const}$, $\omega_1 = 0$, and the unit vectors of the ox_3 axis and radius vector r coincide. Then, $\dot{q} = 0$, the singular numbers of the Hessian U'' relate as $2 : 1 : 1$, and the matrix $q^2 + U''$ is generally

nonsingular unless $|\omega| = \nu$, $|\omega|^2 = \nu^2 + 3\omega_2^2/2$, and $\nu = (\mu/|r|^3)^{1/2}$.

Generalizing these two cases, absolute and relative velocity measurements, we may conclude that the decisive (and, more importantly, physical) condition that ensures the correct statement of the problem and in a certain sense stabilizes its solution is the motion of the point under the action of potential forces with the non-degenerate Hessian. Finally, it is worth noting by way of example that it is this natural condition that provides the asymptotically stable correction of the dynamic operating regime of inertial navigation systems using Doppler-shift-based velocity measurements [4].

REFERENCES

1. A. F. Ioffe, *Fundamental Ideas of Modern Physics* (GITTL, Leningrad, 1949).

2. A. Yu. Ishlinskiĭ, *Classical Mechanics and Inertial Forces* (Nauka, Moscow, 1987).

3. R. Kalman, P. Falb, and M. Arbib, *Topics on Mathematical System Theory* (McGraw-Hill, New York, 1969; Mir, Moscow, 1971).

4. V. D. Andreev, *Theory of Inertial Navigation: Correcting Systems* (Nauka, Moscow, 1967).

Translated by V. Isaakyan

BRIEF COMMUNICATIONS

Water Vapor Condensation and the Collapse of Superheated Water Drops as Possible Causes of Ball Lightning

A. S. Tarnovskiĭ

Samara State Pedagogical University, ul. Gor'kogo 65/67, Samara, 443043 Russia
e-mail: sameta@rambler.ru
Received January 10, 2003

Abstract—Ball lightning is modeled by the bulk of humid air heated to temperatures of 600–650 K and containing a multitude of charged drops and microbubbles of size 10^{-5} cm or less, as well as water vapor heated to a near-critical temperature. The condensation of the vapor in the microbubbles, followed by the cooling of resulting droplets, generates energy spent on the thermal radiation of ball lightning. The radiation of light and radio waves is explained by the motion of ions and electrons in the electric field of charged bubbles and droplets and by the thermal rotational motion of charged droplets. As a result of coagulation, the droplets overheat and tend to collapse. An external electric field, supersaturated water vapor condensation, and a number of other factors may contribute to the ball lightning explosion energy. © *2003 MAIK "Nauka/Interperiodica".*

Ball lightning remains an unresolved phenomenon of nature. Many hypotheses for the origin of ball lightning have been put forward [1–8], such as plasmoids filled with a high-temperature plasma that is confined by the self-magnetic field and various chemical proposals. For example, D. Arago argued that ball lightning consists of nitrogen and ozone oxides impregnated by the "lightning matter." After J. Barry, ball lightning is a hydrocarbon compound that forms in an electric discharge and somehow concentrates in a small volume. There exist hypotheses according to which ball lightning comprises charged particles (Frenkel and Hill): dust or unlike ion clouds separated from each other. Podmoshenskii believes that ball lightning forms from filamentary particles. In an attempt to find the sources of ball lightning energy, exotically minded authors delve into intranuclear processes or even deeper. In 1956, Arabadzhi made a proposal that thunderous electric fields entrap heavy particles, e.g., cosmic rays, into ball lightning, causing a self-sustained nuclear reaction. Note, however, that an elevated radiation level due to the ball lightning effect has not been observed. Vatai [2] believes that the phenomenon of ball lightning proves the existence of new elementary objects: leptonucleons and their clusters. Korshunov [3] advocates that ball lightning appears when a Hooft–Polyakov monopole, which is an efficient catalyst of nucleon decay reactions, passes through the Earth's atmosphere.

Today, the aerogel or cluster model of ball lightning, which has been developed since the late 1970s by Smirnov and his colleagues at the Institute of Thermal Physics (Novosibirsk), is the most popular. According to [4, 5], the active core of ball lightning is a charged structure comprising submicron filaments, that is, a porous fractal cluster with a high chemical capacity. The framework of this structure is almost entirely comprised of empty voids. Energy evolution from a chemically charged fractal cluster may be described in terms of multistep combustion. As an example of such a process and as a ball lightning model, Smirnov suggests the multistep combustion of a fractal cluster of charcoal dust in ozone, which is absorbed by the cluster itself. In this model, the color and glow of ball lightning are produced, as in pyrotechnology, with a composition containing glowing components. The Smirnov model adequately explains various properties of ball lightning. K. Korum and D. Korum created ball lightning in free air [6]. Fireballs observed in their experiments looked like those predicted from the Smirnov fractal model.

However, the properties of ball lightnings discovered in many experiments are so varied and contradictory that a unique theory of lightning origin and behavior can hardly be constructed. In our opinion, several types of ball lightnings are present in nature, each deserving special theoretical consideration. Therefore, various ball lightning models may be to a limited extent adequate. Yet none of more or less realistic models where ball lightning is viewed as an object of volume no more than one cubic meter and density roughly equal to that of air may explain the source of giant energy [4, 5, 7, 8] liberated upon ball lightning explosion. Therefore, together with theoretical models aimed at elucidating the nature of this intriguing phenomenon as such, it is necessary to invoke the models of relevant processes that could help us tackle the question of how "light" and, hence, low-power fireballs may liberate such high energy upon explosion [7, 8].

According to Smirnov [4, 5] and Stakhanov [7], a ball lightning is a red, yellow, white, blue, or much more rarely green glowing spherical object of diameter from 1 cm to 1 m. The mean lifetime is 9 s, and the

mean velocity is 9 m/s. About 51% of ball lightnings come to an end by explosion (sometimes intense and destructive); others die out or decay into parts. The mean energy of a ball lightning is 10 kJ; sometimes, it exceeds 1000 kJ. The thermal radiation is low, 100–500 W, and corresponds to the radiation of a ball 20 to 30 cm in diameter at a temperature 600–650 K [7]. In the optical range, the radiation power is several watts and corresponds to temperatures in the range 2000–10 000 K. This radiation is most likely to be nonstationary [7, p. 151]. It seems that ball lightnings may emit intense nonstationary radiation in the rf range. The electromagnetic radiation associated with thunderstorms and lightnings has a wide frequency spectrum, at least from 10 to 1000 Hz. Note that the high-frequency part of the radiation appears with a delay of 5 ms, i.e., after lightning has been observed [8, p. 215].

As follows from observations, ball lightning is surprisingly stable during its life. Less than 1% of observers indicate a change in its dimensions or color. The observations are different: a lightning may grow, and its color may change from red to violet or vice versa. The only correlation between the physical properties of ball lightning, namely, that its lifetime varies in direct proportion to its dimensions [5], can be explained in simple terms: a larger and brighter object is noticed earlier and seen longer. It was long thought that ball lightnings, like normal linear lightnings, appear during a thunderstorm in the majority of cases. However, relatively new Japanese data based on vast statistics (2060 observations) [5] testify that 89% of ball lightnings are observed in fine weather and only 2.5% in a thunderstorm.

Ball lightning has electrical properties and acts on the human organism like electric current. It may cause dumbness and paresis [5]. The death rate because of ball lightning strokes is 0.5% [7]. The appearance of ball lightning is frequently accompanied by a feeble crack, hissing, or whistle. Sometimes, air smells of sulfur, ozone, or nitrogen oxides. The same smells are sometimes felt under linear lightning discharges and other electrical discharges. In most cases, however, no characteristic signs indicating the presence of any specific ball lightning "material" differing in composition from surrounding air are observed. Sometimes, a vapor cloud or mist forms after ball lightning has exploded or died out.

According to the Smirnov model, the energy accumulated in ball lightning itself is comparable to that given off by striking few matches (several kilojoules). Such energy is too low to explain the destruction of thick tree trunks and massive constructions, damage to water-supply pipes, etc., observed under a stroke. The explosion energy must be as high as several tens of megajoules in these cases. To account for high-power and super-high-power lightning explosions, one must allow for the presence of specific mechanisms that enhance energy evolution upon explosion, whereas

lightning triggers high-power external energy sources. One of these sources is mentioned in [7]. When ball lightning collapses, a conducting channel connecting regions with a large potential drop across them forms, causing a short circuit and releasing a high explosion energy. Stakhanov writes [7]: "The energy accumulates in charged conductors, and ball lightning serves merely as a trigger releasing this energy. However, the possibility of thermal explosions due to the evolution of ball lightning internal energy should not be excluded."

Another mechanism increasing the energy evolution upon ball lightning explosion is suggested in [9]. The explosion of a low-power ball lightning in the space containing supersaturated water vapor causes the explosive, i.e., extremely fast, condensation of moisture on many fragments, including ions from the former lightning. This, in turn, results in a sharp drop in the pressure in the explosion area with the subsequent expansion of nearby humid air strata into this area. This expansion again causes condensation, pressure drop, and further expansion of air strata containing supersaturated water vapor. Thus, a specific chain reaction of successive condensations and expansions of air supersaturated by moisture is initiated. As a result, a cumulative gas-dynamic wave convergent to the center arises. Significant compression work is done by outer "dry" strata of atmospheric air.

An additional energy is liberated because of the moisture condensation. At the final stage, the convergent wave causes a drastic rise in the pressure and temperature in the central area (cumulative effect) and then makes room for a divergent wave. The mechanism suggested may explain intense explosions of ball lightnings and also explosions in upper atmospheric strata [10], including the "mystery of the 20th century"—the explosion above the Podkamennaya Tunguska River in June 1908, which ravaged taiga over an area of more than 2000 km^2 and left no signs of the source of its energy [11]. A feature of a condensation explosion is that it does not require any "explosive" that differs from atmospheric air in composition. The idea that there exist additional mechanisms increasing the explosion energy of ball lightning, in particular, the condensation explosion in the atmosphere due to the decay or explosion of normal low-power ball lightning, does not conflict with any of the realistic ball-lightning models currently available. Conversely, it was noted [12] that the idea of condensation explosion may explain not only the phenomenon of ball lightning but also other atmospheric phenomena. Note that the smell felt after the explosion of ball lightning or after a linear lightning discharge and other electrical discharges may be attributed to the formation of ozone and is certainly not associated with any explosive.

In most cases, however, ball lightning does not "smell" and does not leave any tangible traces behind it. It appears that, along with ball lightnings including foreign (relative to atmospheric air) substances, there exist

all-atmospheric ball lightnings, which consist of heated air and steam, vapor bubbles, and charged water drops. The bubbles and drops may be produced when an electrical discharge is initiated in a humid area. In a number of papers, it was ascertained that nonequilibrium thermal processes may result in a much higher degree of order than was believed earlier [13]. For example, when a high current passes through a fine wire, causing the explosion, explosion fragments are nearly equal in size and have flat edges. Similar effects were observed upon passing a current through liquids [14].

The size of rain drops ranges from 1 μm to several micrometers; that of mist drops, from 0.1 to 1.0 μm [15, p. 220]. The sedimentation rate of such drops is less than 1 cm/h. Let us assume that the discharge of a linear lightning (or any other electrical discharge) generates a vapor–air mixture at a temperature close to the critical temperature $T_c = 374.15°C$ and that this mixture contains a large number of water droplets (or bubbles) that are heated to this temperature and are smaller than mist drops. The moisture content of clouds is usually no more than 10 g/m^3. However, we admit that water droplets in ball lightning have densities, which are roughly equal to, or somewhat greater than, the density of heated humid air in order that, on average, the ball lightning density not exceed the density of the environmental air. Recall that, at near-critical temperatures, the density of moisture-saturated air is less than half the density of air at $T = 20°C$: the density of water and steam at the critical temperature is $\rho_c = 0.307$ g/cm^3. According to our estimates, a medium-size ball lightning occupying a volume of about 10 dm^3 contains almost 10 g of water and steam, which may release an energy of about 40 kJ upon cooling. Such a ball lightning may emit thermal radiation of power 1–2 kW for 10 s without changing its parameters. If the droplets are sufficiently small, on the order of 10^{-6}–10^{-7} cm across, their surface energy may contribute significantly to the ball lightning energy. The total surface energy of water, 118 mJ/m^3, is temperature independent [15, p. 36] and may be released as heat upon drop coagulation.

Consider a water bubble of radius r and density 0.45 g/cm^3 at a temperature $T \geq 300°C$ that contains saturated steam under a pressure p that is much higher than the atmospheric pressure. The overpressure is balanced by the surface tension of the water film:

$$p = \alpha/r. \qquad (1)$$

In (1), the pressure and the surface tension coefficient α depend on temperature. With $T = 300°C$, $\alpha = 15$ erg/cm, and $p = 90$ kg/cm$^2 = 10^8$ dyn/cm^2 [15, p. 293]. Hence, $r \sim 10^{-7}$ cm. With $T = 200°C$, $r \leq 10^{-5}$ cm. Such water bubbles are temperature stable: as the temperature declines, the surface tension coefficient α grows. Accordingly, the pressure of the water film increases and the equilibrium state recovers owing to the compression force. And conversely, as the temperature spontaneously increases, the surface tension

declines, the steam expands, the temperature drops, and the equilibrium state recovers. Therefore, heat evolution from such bubbles due to the condensation of the steam inside is a stable and quasi-equilibrium process. Eventually, the bubbles shrink and the inner pressure and temperature increase.

This process lasts until the entire steam inside the bubbles condenses completely. Then, the coagulation of the resulting water drops, which are heated to a near-critical temperature, and their cooling comes into play. If the coagulation is fast, the water in the drops becomes overheated and the drops tend to collapse. When exploding, such ball lightning liberates a relatively low energy. However, the explosion disperses many fragments of the former ball lightning in the environment, including droplets and microbubbles. This process may initiate a secondary explosion of much higher power, e.g., because of the fast condensation of supersaturated water vapor in the environment, short-circuit effect, etc. If the density of resulting hot water drops is insufficient for the fast coagulation or the drops bear large like charges, hindering the coagulation, they slowly cool down, evaporate, or spread in the environment. Such ball lightning dies out noiselessly, sometimes decaying into fragments of various size.

Certainly, the formation of bubbles is not an obligatory stage of the ball lightning's existence. A high-power electrical discharge may immediately produce superheated or merely hot drops of any size that are suspended in hot and humid air.

To adequately explain the electrical properties of a ball lightning and shed light on the nature of emitted electromagnetic radiation, one must admit that at least some of the drops are charged. The radiation may be emitted by ions bearing a charge of the same sign as the drops. Having separated from a parent drop, the ion comes into interaction with it and gain a kinetic energy $E = Ze^2 r^{-1}$, which corresponds to a temperature $T \sim 10^4$ K with $r \sim 10^{-7}$ cm and $Z = 1$. In subsequent collisions, the ions will emit visible radiation under nonequilibrium conditions. Coarser drops with a higher charge for which the product $Ze^2 r^{-1}$ is of the same order of magnitude will also emit radiation in the visible range. Drops for which $Ze^2 r^{-1}$ is sufficiently large or small will emit radiation in the UV or IR range, respectively. Ions and electrons opposite in sign to the drop, when approaching it, also will gain an energy sufficient to emit radiation in the visible or invisible range of the spectrum.

The fact that ball lightnings are colored green very seldom, if at all, may be explained on a qualitative basis. The point is that the color green corresponds to the center of the visible range. Let us take into account (or assume) that the effective emission spectrum of ball lightning is rather wide and does not have a distinct maximum at its center. Then, the radiation of ball lightning for which the color green lies at the center of the emission spectrum will be perceived by an observer as

white light. If the center of the spectrum is shifted away toward the violet or red edge of the visible range, the radiation will be appropriately colored. For both the equilibrium and nonequilibrium radiation of ball lightning, the probability that the emission spectrum is sufficiently narrow with green at the center is small. That is why green ball lightnings are rare in occurrence.

The radio radiation of ball lightning may be related to the motion of charged droplets. After the vapor inside the microbubbles has condensed and thermal relaxation has come to end, the energy per degree of freedom (including rotational degree of freedom) of a droplet will roughly equal kT. Such droplets will rotate with an angular velocity ω such that

$$Mr^2\omega^2 \sim kT. \qquad (2)$$

Putting $M \sim r^3 \approx 10^{-21}$ g (M is the mass of a droplet) and $T \approx 600$ K, we find that the radiation has a frequency $\omega \sim 10^{10}$ s^{-1}, which corresponds to the centimeter range. Drops of size $r \sim 10^{-5}$ cm will emit with a frequency $\omega \sim 10^{-6}$ s^{-1}, i.e., in the mid- and long-rf range. What is the reason for a delay in the rf radiation from ball lightning? At the initial time instant, lightning consists mainly of water–vapor bubbles, which do not emit in the rf range. After the condensation of the vapor contained in the microbubbles, the resulting droplets also radiate mostly in the optical, rather than in the radio-frequency, range. Only charged coarse drops formed by coagulation begin to emit in the rf range. The formation of sufficiently large charged drops from microbubbles takes some time. This reasoning explains, at least qualitatively, why the rf radiation lags behind by about 5 ms.

Consider the stability of ball lightning. For spherical dispersed particles bearing a minor electrical charge, the total energy vs. distance h dependence has a maximum and two minima (see figure) provided that their radius exceeds the diffusion layer thickness; i.e., $r \geq 10^{-7}$ cm [15, p. 281]. The first minimum corresponds to particle (droplet) coagulation, whereas the other reflects molecular attraction. For not very large particle spacings, electrostatic repulsion prevents their approach. The height $U(h)$ of the maximum is a potential barrier that defines the coagulation probability. The
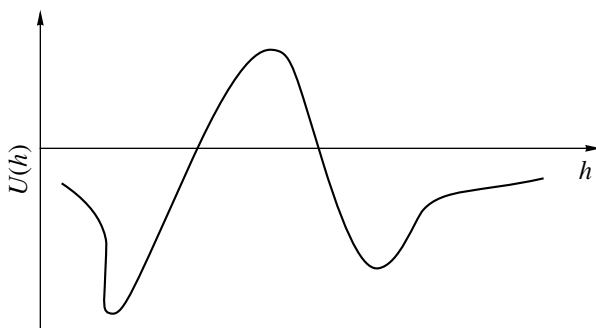
mutual attraction of widely spaced drops generates surface forces at the boundary of ball lightning, causing its stability. The surface tension at the boundary of lightning is obviously much lower than the surface tension at the drop–air interface (at least by two orders of magnitude). However, even a low surface tension (on the order of 10^{-7} J/cm^2) makes a ball lightning stable if its density differs from that of environmental air by no more than 1% [7, p. 119].

For a solitary charged drop to be stable, it is necessary that its electrostatic energy be lower than the surface energy of attraction; that is,

$$Z^2 e^2 r^{-1} < 4\pi r^2 \sigma. \qquad (3)$$

If a drop arises through the coagulation of several droplets with equal volumes and charges, its charge will be proportional to its volume: $Z \sim r^3$. Consequently, the electrostatic energy of the drop during coagulation increases as r^5, while its molecular surface energy varies as r^2. Therefore, as the size of charged droplets grows, their coagulation breaks condition (3) and the resulting drop becomes unstable. On the other hand, a higher charge of the resulting drop prevents further coagulation. Thus, the relatively stable state where droplets have some intermediate sizes and charges occurs. The situation where charged water droplets lose their electrical charge is much more dangerous to the existence of ball lightning. In this situation, the potential barrier disappears and slow coagulation changes to fast coagulation [15] with the characteristic time $\theta = \eta/2kT\nu_0$, where η is the viscosity of the medium and ν_0 is the number of particles per unit volume. For $T = 600$ K, the air viscosity is $\eta = 3 \times 10^{-4}$ g/(cm s). If particles have a size of 10^{-7} cm, a mass of 10^{-3} g per cubic centimeter, and a density of 1, then we have $\nu_0 \cong 10^{16}$ cm^{-3}. Then, the coagulation half-time is $\theta = 10^{-8}$ s. After a time τ, the number of particles per unit volume will be

$$\nu = \frac{\nu_0}{1 + \tau/\theta}.$$

By the time $\tau = 10$ s, the number of droplets will decrease 10^9 times and their size will reach 10^{-4} cm.

These particles are optically active: they reflect and refract light. If the number of such droplets in ball lightning is appreciable, the presence of external light sources will cause typical optical effects. External light reflected and refracted by droplets adds up to the glow of the remaining charged droplets and is perceived as the self-glow of ball lightning. The apparent glow of the lightning will for some time remain bright and at the same time nonuniform over the volume because of the random motion of coarse drops relative to each other and their coagulation. The glow varies with time and appears iridescent. Mechanical processes attendant on the existence of such a "water" lightning, the coagulation of weakly charged droplets and the decay of

Figure.

coarser charged drops, may be responsible for the acoustic effects observed, such as crack, hissing, or whistle.

We have considered the case of extremely fine droplets. However, ball lightning may incorporate drops of various sizes. Charged droplets provide the long lifetime, stability, and glow of ball lightning. Coarser drops, 10^{-4} cm or more in size, start fast coagulating at a certain time instant, cutting the lifetime of the lightning. If the droplets had a chance to cool down by the instant they coagulated to form coarse drops, the lightning will noiselessly die out or decay. If at least some of the droplets had coarsen to a size $r \geq 10^{-4}$ cm via fast coagulation before they cooled below the boiling point, water in such large drops overheats and the drops become dangerously explosive (i.e., tend to collapse). As is known [15, p. 189], the overheat of a pure liquid may be rather high for a short time.

Water drops heated to 200°C are certainly unstable. The spontaneous boiling of one such drop will break it into many fine fragments. Falling into neighboring overheated drops, these fragments will initiate a chain reaction of boilings and explosions of overheated drops. Such is, in our opinion, a probable explosion scenario for one possible type of ball lightning, namely, ball lightning that leaves behind no signs of its appearance and disappearance other than an additional amount of moisture. Of course, this extra moisture in the air cannot in itself attract researchers' attention. Air consists of many components: nitrogen, oxygen, argon, etc. However, these components are present in fixed proportions and any local change in their relative concentration cannot go unnoticed. As to moisture, its concentration in air may vary arbitrarily over wide limits; therefore, a change in the moisture content does not attract the attention of observers. That is why the hypothesis about "all-water" ball lightning seems to be promising and has a right to exist.

REFERENCES

1. Izv. Vyssh. Uchebn. Zaved. Fiz., No. 3 (1992).
2. E. Vatai, Nuovo Cimento A **101**, 905 (1989).
3. V. K. Korshunov, Int. J. Mod. Phys. A **5**, 1629 (1990).
4. B. M. Smirnov, *Problem of Ball Lightning* (Nauka, Moscow, 1988).
5. B. M. Smirnov, Usp. Fiz. Nauk **160** (4), 1 (1990) [Sov. Phys. Usp. **33**, 261 (1990)].
6. K. L. Korum and D. R. Korum, Usp. Fiz. Nauk **160** (4), 47 (1990).
7. I. P. Stakhanov, *On the Physical Nature of Ball Lightning* (Énergoatomizdat, Moscow, 1985).
8. S. Singer, *Nature of Ball Lightning* (Plenum, New York, 1971; Mir, Moscow, 1973).
9. A. S. Tarnovskiĭ, Zh. Tekh. Fiz. **60**, 200 (1990) [Sov. Phys. Tech. Phys. **35**, 395 (1990)].
10. M. T. Dmitriev, Aviatsiya i Kosmonavtika, No. 3, 44 (1979).
11. A. S. Tarnovskiĭ, Aviatsiya i Kosmonavtika, No. 7, 38 (1992).
12. B. M. Smirnov, Phys. Rep., Phys. Lett. **224**, 151 (1993).
13. S. A. Maĭorov, A. I. Tkachev, and S. I. Yakovenko, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 11, 3 (1991).
14. H. Apsden, Phys. Lett. A **111**, 22 (1985).
15. Yu. G. Frolov, *Course of Colloid Chemistry: Surface Phenomena and Disperse Systems* (Khimiya, Moscow, 1989).

Translated by V. Isaakyan

BRIEF
COMMUNICATIONS

Core–Shell Structure in Fullerene-Containing Molecular Complexes

V. A. Reznikov

St.Petersburg State University, Universitetskaya nab. 7/9, St.Petersburg, 199164 Russia
Received March 4, 2003

Abstract—A shell model that describes the formation of molecular complexes around a C_{60} molecule upon the solid-phase interaction of the powders is considered. From the absorption spectra, it follows that a reacting particle is a fullerene dimer. © *2003 MAIK "Nauka/Interperiodica".*

The insignificant asymmetry of C_{60} molecules entering into the composition of a number of complex compounds produced by the mechanical interaction of powders suggests a fullerene-like structure of functional molecular groups on the van der Waals surface of C_{60}. In C_{60}–KHal complexes (Hal stands for halogen), adsorption interaction between the components is responsible for the vibrational spectra of C_{60}. The fractal coral-like pattern of crystallization from a saltwater–C_{60}–KHal solution [1] suggests that molecular complexes (MCs) serve as seeds and represent core–shell structures. The effect of a halogen entering into an MC on the C_{60} symmetry and the reduction of this effect when the MC is adsorbed in the volume of amorphized KHal [2] or is dissolved in a polar solvent [1] means that the formation of the second and subsequent coordination spheres decreases the potential gradient on the surface of a carbon cell. In particular, the thermal stability of C_{60}–KJ MCs and the formation of octahedral crystals from them on pressing the powders are indicative of a high degree of symmetry in these MCs and a slight change in the lattice constant of KJ incorporated into the shell cluster [3]. On the contrary, the decomposition of C_{60}–KCl(NaCl) MCs at 120–160°C testifies to the amorphous state of the shell.

Additional information on the structure of C_{60}–KHal MCs can be extracted from the absorption spectra of their solutions.

Figure 1 shows the absorption spectra of a KJ–C_{60} MC in a HCl solution. The spectra were recorded (*A*) before and (*B*) after the irradiation of the solution in the range 3.0–3.5 eV. For other alkali halides, the absorption spectra of associated MCs dissolved in HCl differ insignificantly. For comparison, Fig. 1 shows the absorption spectrum (curve *C*) of the C_{60}–KCl MC dissolved in a nonpolar (hydrochloric) solvent. The table lists the peak energies of the bands in the spectra of MC solutions with various halogens, the shifts of the peaks with respect to the reference values for the C_{60} spectra in nonpolar solvents, and calculated values of the π

band for unsolvated C_{60} [4, 5]. The UV band maxima were measured accurate to 1.5–2.0 meV.

From the table it is seen that, in all cases (except for the as-prepared C_{60}–KCl solution), the band maxima are shifted by the characteristic vibrational states of C_{60}. The pairs of bands at 3.475–3.95 and 5.34–3.913 eV can be considered as the result of splitting the π states of solvated C_{60} (the bands with the maxima at 3.712 and 3.626 eV that are shifted from the calculated value 3.81 eV [4] by 776 and 1460 cm^{-1}).

The split of surface plasmons is observed for ultradisperse metal particles of ellipsoidal shape [6]. Similar high-energy components of the π band in the spectrum of highly symmetric C_{60}–$H_2 TPP$–C_{60} porphyrin–fullerene complexes [7] (*TPP* is tetraphenyl porphyrin) were explained by Coulomb interaction between the polarized π states of C_{60} and central H^+ ions in the porphyrin macrocycle. The dependence of the shift of the π bands on the type of a halogen entering into the MC composition (≈0.2 and 0.24 meV) leads us to assume that Br^0 and J^0 located between two C_{60}
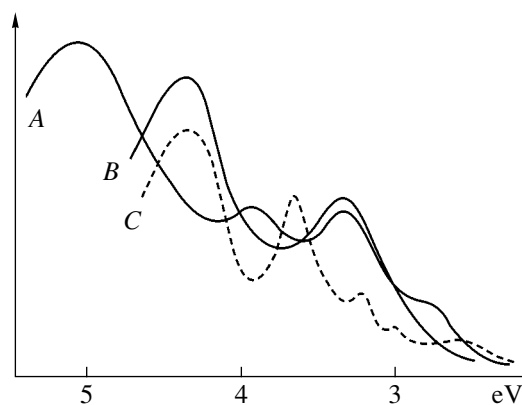

Fig. 1. Absorption spectra of the KJ–C_{60} molecular complexes in the HCl solution (*A*) before and (*B*) after optical irradiation; (*C*) refers to the C_{60}–KCl MC in the nonpolar solvent.

Table

Composition	Band maxima, eV				Shifts of the maxima, cm^{-1}
KJ–C$_{60}$	5.025	3.913	3.34	2.8	1460, 776, 496, 172
$h\nu$	4.592	–	3.4	–	1570, 272
KBr–C$_{60}$	5.063	3.95	3.475	2.83	1460, 496
$h\nu$	4.625	–	3.415	–	1570, 172
KCl–C$_{60}$	4.976	–	3.375	2.75	1570, 776, 1230
$h\nu$	5.04	–	3.435	–	1570, 172
KCl–C$_{60}$ Incomplete solubility	4.53	3.71	3.24	3.024	1570, 776, 172
C$_{60}$	4.82	3.71	3.435	3.045	

atoms act as an h^+ center. In such a model, the absence of the high-energy component of the π band in the spectrum of a chlorine-shell MC dissolved in HCl may be related to the relatively high electron affinity $E_a\mathrm{Cl}^0 = 3.61$ eV and/or to a different shell structure.

The low-energy bands are shifted with respect to the characteristic band at 3.045 eV, which is associated with the electron transition in a resonantly excited center involved in a C$_{60}$ dimer [4, 5]. The optical quenching of the π band high-energy components and low-energy bands means the decay of the dimeric MC. The bathochromic (long-wavelength) shift of the π–σ bands by the vibrational states of C$_{60}$ in the spectra of the C$_{60}$–KJ and C$_{60}$–KBr MCs corresponds to the excitation of double bonds in monomolecular C$_{60}$ in the presence of the ligand. The hypsochromic (short-wave) shift by 496 cm^{-1} in the spectrum of the irradiated C$_{60}$–KCl solution (in particular, the coincidence of the π band maximum with the calculated value for unsolvated C$_{60}$) indicates that the symmetry of the shell rises toward the fullerene-like distribution of the electron density with a simultaneous increase in the binding energy (E_b) of the MC components.

The shift of the π band low-energy components by the characteristic vibrational states of C$_{60}$ with respect to the reference values of the solvated molecule and to the calculated value for the same molecule in the unsolvated state suggests a mixed type of shell. Similarly, the correlation between the high-energy components of the π band and the solvated state of C$_{60}$, as well as the larger half-width of the low-energy bands compared to the vibronic bands at 3.024 and 3.01 eV in the spectra of the same MCs dissolved in the nonpolar solvent (Fig. 1, curve C), indicates that the components of the polar solvent enter into the composition of the shell but disappear upon the photoinduced raising of the C$_{60}$ symmetry. The bathochromic shift of the band at 3.045 eV by the vibrational states of C$_{60}$ is observed in solvents with polar (or readily polarizable) groups (methylpyrrolidone and carbon disulfide). According to

Fig. 1 (curve C), a similar shift of the band at 3.045 eV is also possible for mixed-type shells. The broadening of the low-energy bands by 0.10–0.11 eV reflects the incorporation of H$^+$ ions into the alkali-halide shell and the Coulomb interaction of the H$^+$ ions with the polarized π and π–σ states of C$_{60}$.

The maxima of the π–σ bands and those of the corresponding low-energy bands (see table) are shifted by 2.225 eV. At the same time, the low-energy bands are equidistant from the bands of unsolvated C$_{60}$: 3.34 ($\omega = 776$ cm^{-1}), 3.314 ($\omega = 172$ cm^{-1}), and 3.24 eV ($\omega = 1570$ cm^{-1}), and from the virtual band 2.26 eV (2.225 + 0.034 eV). In the absorption spectrum of the dissolved C$_{60}$–H$_2$TPP–C$_{60}$ MC solution, the band 2.26 eV is due to the intracomplex transition between the low-energy level of central delocalized H$^+$ ions and collective π states [7], which corresponds to E_aH$^+$. Thus, the shift of the high-energy band maxima by 1.113 eV (see table) characterizes the interrelated transitions in the field of H$^+$ ions delocalized between the polarized π and π–σ states of C$_{60}$. The superposition of the MC excited states at the energy levels E_b2C$_{60}$ = 3.045 eV and E_aH$^+$ = 2.26 eV corresponds to E_aC$_{60}$ = 2.65 eV with an accuracy of ≈ 2.0 meV. Consequently, in the acidic medium, the equilibrium state of H$^+$ ions between two C$_{60}$ in the MC is possible. In the presence of H$^+$ ions, the values E_aJ$_2$(Br$_2$) = 2.55 eV [8] must increase to E_aC$_{60}$ or E_aC$_{70}$ = 2.69 eV. At these values, the potential in a pair MC reaches a minimum when the symmetry of the carbon cell is unbroken.

The energy of interaction between the C$_{60}$–H$_2$TPP–C$_{60}$ components was evaluated as 153 meV and coincided with the hypsochromic shift of the π band [7]. It is reasonable to assume that the shift of the π band high-energy components in the spectra of the C$_{60}$–KHal MCs dissolved in HCl (by 0.20–0.24 eV) also corresponds to the interaction energy of the MC components. For iodide-containing MCs, the ratio of the anticipated value $E_b = 203$ meV to that in the porphyrin–fullerene
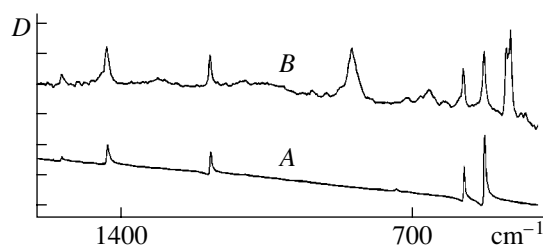
Fig. 2. IR absorption spectra of C_{60} incorporated into the (*A*) KBr matrix and (*B*) sulfur matrix.

MC coincides with the ratio $E_aJ^0/E_aH^+ = 1.326$. The decrease in the ratio of the same energy parameters for bromide-containing MCs by a factor of 1.04 is most likely to be related to a decrease in the lattice constant of KBr incorporated into the complex. In the MCs with the chloride shell, the shift of the π–σ band, as well as of the virtual high-energy component of the π band when its maximum is shifted by 1.13 eV with respect to the low-energy band (2.75 eV), coincides with the binding energy in a porphyrin–fullerene complex. This means that the shell combined with the solvent forms a unified system of charge carriers, that is, is virtually a solvation sheath.

In view of lattice constant values of KHal, the van der Waals surface of C_{60} can be covered by 12 KJ (or KBr) molecules in the form of icosahedral clusters and 16 KCl molecules. In such a model, the decrease in the ionic radius ratio from KJ to KBr (as well as the decrease in the lattice constants) is responsible for the uniform compression of C_{60} in the bromide-containing shell [2]. In chloride-shell MCs, the symmetry of the carbon cell persists when 32 Cl atoms are arranged over the planes of cyclic C–C bonds. Consequently, in the solid phase, the shell is bilayer, whereas in the HCl solution, it may be monolayer, consisting of Cl_2 molecules statistically dissociated in the field of H^+ ions.

To validate the dimeric model of the MCs, we also considered solid-phase interaction between C_{60} fullerite and stick sulfur. Compared to the IR spectrum of C_{60} in the KHal matrix, the IR spectrum of C_{60} in the sulfur matrix (Fig. 2, curve *B*) is characterized by the broader bands and decreased relative intensities of the bands at 525 and 575 cm^{-1}. Sulfur gives the intense bands of S_2 (with maxima at 462 and 471 cm^{-1}) and S_2^+ (836 cm^{-1}) [8].

The presence of the S_2 bands in the spectrum testifies to the interaction between S_8 and C_{60}, which does not break the symmetry of C_{60}. This is possible when a S^+–S^- network forms. A sphere of radius 5.01 Å can accomodate five deformed S_8 rings. To form a fullerene-like shell, 44 atoms are needed, whereas an icosahedral cluster requires 42 atoms (at a S–S spacing of 2.05 Å). The intermolecular spacing in C_{60} fullerite correlate with the S_2 molecule length; therefore, the

surface of a C_{60} dimer accomodates ten S_8 molecules. In a $40S_2$–$2C_{60}$ MC, the ratio C/S satisfies the stoichiometric composition of a C_3S_2 quasi-drop, where a decrease in the density ($\rho = 1.27$ g/cm^3) by a factor of 1.5 with respect to the sulfur density is due to the larger intermolecular spacing and the presence of free space in C_{60}. In C_3S_2 MCs, the potential reaches a minimum at the vibronic level of unsolvated C_{60}:

$$\bar{E}(E_bS^0, E_a^0 \quad \text{and} \quad E_aC_{60}) = 3.401 \text{ eV},$$

$$\bar{E}(E_bS_2^-, E_aS^0) = 3.013 \text{ eV} \ (\omega = 272 \text{ cm}^{-1}),$$

where $E_bS^0 = 6.23$ eV [9], $E_aS^0 = 2.077$ eV, and $E_bS_2^- = 3.95$ eV [8].

The Coulomb interaction of the set of C^+ ions in a carbon cell with S_2^- dimers is in accordance with the CS_2 band resolution (at 657 cm^{-1}).

For a C_{60} dimer as an elementary reacting particle, the chemical restrictions imposed on a nonpolar C_{60} molecule are lifted. A C_{60} dimer is stable because of π–π exchange interaction and can be considered as a basic repeating unity of fullerite [4]. The formation of a core–shell MC results from the energy organization of homogeneous or heterogeneous molecular clusters on the surface of a flexible-geometry sorbent, which in full measure holds true for solid-phase fullerene–polymer interaction [5].

REFERENCES

1. V. A. Reznikov and A. A. Sukhanov, Pis'ma Zh. Tekh. Fiz. **25** (8), 45 (1999) [Tech. Phys. Lett. **25**, 313 (1999)].
2. A. A. Sukhanov and V. A. Reznikov, Pis'ma Zh. Tekh. Fiz. **25** (9), 56 (1999) [Tech. Phys. Lett. **25**, 363 (1999)].
3. G. É. Yalovega, A. V. Soldatov, K. Novak, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **42**, 1889 (2000) [Phys. Solid State **42**, 1942 (2000)].
4. V. A. Reznikov and Yu. S. Polekhovskiĭ, Pis'ma Zh. Tekh. Fiz. **26** (15), 94 (2000) [Tech. Phys. Lett. **26**, 689 (2000)].
5. V. A. Reznikov, E. Yu. Melenevskaya, and L. S. Litvinova, Vysokomol. Soedin. **42**, 229 (2000).
6. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983; Mir, Moscow, 1986).
7. E. Yu. Melenevskaya, V. A. Reznikov, and L. S. Litvinova, Vysokomol. Soedin. **41**, 864 (1999).
8. B. M. Smirnov and A. S. Yatsenko, Usp. Fiz. Nauk **166**, 230 (1996).
9. S. A. Shchukarev, *Inorganic Chemistry* (Vysshaya Shkola, Moscow, 1974), Vol. 2.

Translated by Yu. Vishnyakov

BRIEF
COMMUNICATIONS

Formation of Polarity-Sensitive Input *I–V* Characteristics of a Bidirectional Triode Thyristor

S. B. Baklanov, N. T. Gurin, E. V. Lychagin, S. G. Novikov, A. V. Kartavenko, and M. A. Kostylov

Ul'yanovsk State University, Ul'yanovsk, 432970 Russia
e-mail: soy@sv.uven.ru
Received March 11, 2003

Abstract—A family of input *I–V* characteristics of a planar diffused bidirectional triode thyristor (triac) is obtained. The formation of an *N*-shaped portion in the input characteristic is explained. It is shown that the input *N*-shaped characteristic may form in various input circuits of a triac depending on the voltage polarity across power electrodes. © *2003 MAIK "Nauka/Interperiodica"*.

Multilayer semiconductor devices such as triacs are, in essence, elementary functional devices because of the presence of internal positive feedback. This feature allows designers to form negative-differential-resistance (NDR) portions in the *I–V* characteristics. The search for new applications of triac structures is continuing. In [1], triacs were tested as elements of optrons. Ways of improving the magnetic sensitivity of the structure were suggested in [2]. Current studies of triacs are aimed, as a rule, at optimizing the design and process parameters and at improving the standard *S*-shaped input *I–V* characteristic. However, a triac as a device with embedded positive feedback exhibits the duality of input and output characteristics. That is, an *S*-shaped *I–V* curve single-valued in current is dual to an *N*-shaped input characteristic single-valued in voltage; they may be brought into coincidence by reversing the current and voltage axes [3]. Note that the formation mechanism of the *N*-shaped input characteristic that was suggested in [4] contradicts available experimental data.

To refine the formation mechanism of the input characteristic, we studied a planar diffused triac built around an integrated module of two coplanar p_1–n_0–p_2–n_1 thyristors. The thyristors were made in the bulk of the semiconductor n_0 and connected inversely parallel to each other through an external wiring (the p_1 region of one thyristor was connected to the n_1 region of the other) [2]. The circuit consisting of the gate electrode (p_2 region) and nearby power electrode (n_1 region) forms the input; the output is the circuit made up of the power electrodes (p_1 and n_1 regions). Since the triac under study consists of direct and inverse thyristors, we will analyze the physics and operation of only one of them. The *N*-shaped input characteristic with an NDR portion (Fig. 1b) arises in the cathode–gate circuit. In this circuit, the current *I'* is a function of the voltage *U'*, while the voltage *U* between the anode and cathode

serves as a control parameter. In the output *S*-shaped characteristic (Fig. 1a), the voltage *U* is a function of the output current *I*, while the current *I'* is a control current.

Consider the formation of various parts of the input characteristic in relation to the output characteristic. Let the initial output characteristic be schematically represented by curve *1* in Fig. 1a, where *AB* is the load curve corresponding to the bistable operation of the
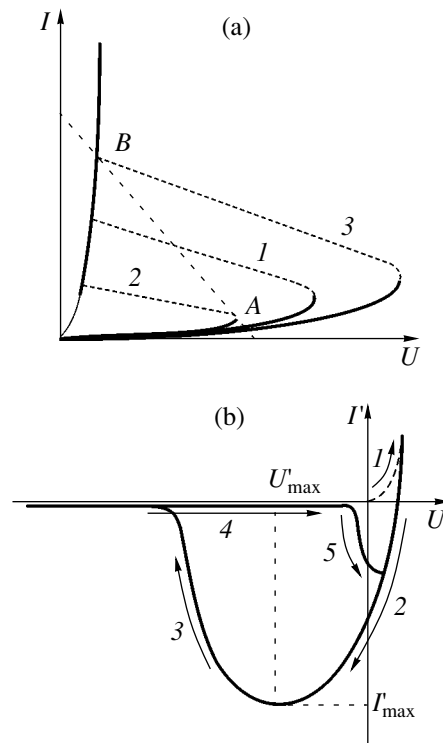

Fig. 1. Schematic representations of the (a) output and (b) input characteristics of a triac.
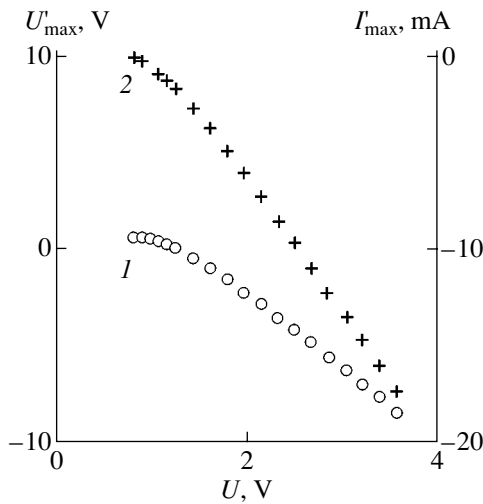
Fig. 2. Experimental dependences of the (*1*) supply voltage $U'_{\max}$ and (*2*) peak current $I'_{\max}$ on the control voltage U.

device. With a positive current I' applied to the gate electrode, curve *1* passes to curve *2* and the thyristor structure switches from the off to on (high-conductivity) state (from point *A* to point *B* along the load straight line). In the input curve (Fig. 1b), the working point moves in the direction shown by arrow *1*. In this state, all holes injected from the anodic p_1 region to the base n_0 reach the base p_2. The mismatch between the value of the positive voltage U' on the p_2–n_1 control junction and the majority carrier concentration in the base p_2 generated a current that is negatively directed in the given case. Simultaneously, the input characteristic exhibits a slight decrease in the forward current and a peak. When the current I' through the control electrode decreases, curve *2* of the output characteristic (Fig. 1a) passes to curve *1*. However, because of the bistable operating mode, the thyristor remains on, since the current through point *B* exceeds the holding current. This causes the working point to move in the direction of arrow *2* in the input characteristic (Fig. 1b) and, accordingly, causes the negative current to grow, since the negative polarity of the voltage U' on the p_2–n_1 junction favors the drainage of holes. However, as the control current I' decreases further, the output characteristic achieves the state (Fig. 1a, curve *3*) where the holding current becomes greater than the current at working point *B*. In this situation, the negative current of the input characteristic reaches the maximal value $I'_{\max}$. With the current I' through the control electrode decreasing still further, the thyristor structure switches off (its state changes from point *B* to point *A* along the load straight line *A*; Fig. 1a). The transfer of holes is sharply reduced, and the negative current in the input characteristic decreases (in the direction shown by arrow *3* in Fig. 1b) to the saturation current of a reverse-biased p–n junction.

The return to positive values of the voltage U' across the p_2–n_1 control junction follows arrow *4* (by analogy with the travel of the working point in the output characteristic upon switching off, when the back and forward currents of the working point in the input characteristic do not coincide). Then, the structure switches on again, and the negative current I' simultaneously increases (Fig. 1b, arrow *5*).

Experimental data for the basic parameters of the triac's output characteristic at 20°C are shown in Fig. 2. The threshold value of the control voltage U at which the NDR portion appears was found to be 0.75 V. The peak reaches the breakdown range at the control voltage $U = 3.6$ V. In this case, the maximum in voltage changes from 0.75 to –8.8 V and in current from 0 to –18 mA. The output characteristic is readily controllable: the curves $U'_{\max}(U)$ and $I'_{\max}(U)$ are almost linear in most of the control voltage range.

Similar measurements taken from the second thyristor incorporated into the triac gave identical results. It should be noted that the *N*-shaped characteristics in the input circuits of both thyristors cannot be observed simultaneously, since the control voltage polarity makes possible the occurrence of this characteristic only in one input circuit. Therefore, triacs seem to be promising as voltage-polarity-sensitive devices: different pairs of input electrodes may exhibit the *N*-shaped *I–V* characteristic depending on the voltage polarity on power electrodes.

Thus, we have shown that the formation of the *N*-shaped portion in the input characteristics of a triac (or in the input characteristic of either of the thyristors) is explained by the unique correspondence between the input and output *I–V* characteristics. This portion results from the presence of the on and off states in the *S*-shaped output characteristic, which differ in their degree of charge carrier transfer. The *N*-shaped characteristic appears across one or the other pair of input electrodes according to the polarity of the voltage applied to the power electrodes.

REFERENCES

1. S. B. Baklanov, N. T. Gurin, and S. G. Novikov, Izv. Vyssh. Uchebn. Zaved. Élektronika, No. 6, 49 (1997).

2. S. I. Vorontsov, S. B. Baklanov, N. T. Gurin, and S. G. Novikov, Pis'ma Zh. Tekh. Fiz. **28** (19), 37 (2002) [Tech. Phys. Lett. **28**, 812 (2002)].

3. S. A. Garyainov and I. D. Abezgauz, *Negative-Resistance Semiconductor Devices* (Énergiya, Moscow, 1970).

4. A. A. Kolosov, Yu. I. Gorbunov, and Yu. E. Naumov, *Solid-State Semiconductor Circuits* (Sov. Radio, Moscow, 1965).

Translated by V. Isaakyan