# Photophoresis of coarse aerosol particles with nonuniform thermal conductivity

Yu. I. Yalamov and A. S. Khasanov

*Moscow Pedagogic University, 107005 Moscow, Russia*

(Submitted November 10, 1996)
Zh. Tekh. Fiz. **68**, 1–6 (April 1998)

The photophoresis of a coarse solid spherical aerosol particle in a one-component gas of nonuniform temperature is examined with consideration of the inertial effects in the hydrodynamic equations and the temperature jump in the Knudsen layer. The problem is solved in the spherical coordinates $r$, $\Theta$, and $\varphi$. The photophoresis of a homogeneous particle is considered first. Then the results are generalized to an inhomogeneous particle. A particle whose thermal conductivity $\varkappa_i$ varies as a function of $r$ is chosen as a model which describes a broad class of natural and artificially produced aerosol particles. It is shown that the error can be significant if the variable internal thermal conductivity $\varkappa_i = \varkappa_i(r)$ of the particle is ignored and only the value of the thermal conductivity on its surface $\varkappa_i(a)$ is considered, on the assumption that the particle is homogeneous. It is also shown that a particle with a variable internal thermal conductivity $\varkappa_i = \varkappa_i(r)$ and a density of heat sources within it $q_i(r,\Theta)$ can be regarded as a homogeneous particle with a thermal conductivity $\gamma\varkappa_i(a)$ and a heat-source density $m(r)q_i(r,\Theta)$. Recurrence formulas for $\gamma$ and $m(r)$ in the general case are presented. Analytical expressions for $\gamma$ and $m(r)$ are found for a model particle with pronounced inhomogeneity. © *1998 American Institute of Physics.* [S1063-7842(98)00104-4]

## INTRODUCTION

The idea that a particle can move under the influence of light was suggested long ago by Kepler. Newton's corpuscular theory reinforced this idea, and the existence of light pressure was demonstrated in Russia by Lebedev. However, Ehrenhaft[1] discovered an effect in the motion of dust particles suspended in air in the beam of a high-intensity lamp: some particles moved toward the light source. This effect could not be attributed to the light pressure force. Ehrenhaft called the effect which he discovered photophoresis. The motion of particles in the direction of light propagation was termed positive photophoresis, and motion in the opposite direction was termed negative photophoresis. This effect can be explained briefly as follows. The absorption of light by a particle leads to distribution of the electromagnetic energy of the incident optical radiation throughout the volume of the particle. Sources of thermal energy appear within the particle with a certain volume density $q_i(r,\Theta)$ and heat it nonuniformly. Gas molecules colliding with the surface of the particle are reflected with a greater velocity from the heated side of the particle than from the cold side. As a result, the particle acquires an uncompensated momentum directed from the warm side of the particle to its cold side. Either the illuminated or the dark side of the particle can be warmer, depending on the dimensions and the optical properties of the particle material. Therefore, both positive and negative photophoresis can occur. In addition, if the radiation flux is nonuniform over its cross section, transverse motion of the particle relative to the direction of propagation of the electromagnetic radiation can appear in a gas.[2]

After Ehrenhaft, the effect was investigated in a number of studies, but the motion of particles in an optical radiation field was only of scientific interest at first, since few manifestations of this effect of practical importance were noted (for example, the influence of solar radiation on the settling of aerosol particles in the earth's atmosphere was considered in Refs. 3 and 4). In recent years the situation has changed dramatically as a result of the use of lasers, and the interest in photophoresis has increased. There have been a number of theoretical and theoretical papers on the theory of photophoresis (see, for example, Refs. 5–8). Numerous applications of the motion of macroparticles in a laser beam have been proposed: the separation of particles in a liquid, the optical levitation of particles in air and in a vacuum, the trapping and retention of particles in a laser beam, etc. The high monochromaticity of laser radiation and the possibility of tuning the wavelength permit the control of the motion of macroparticles and the selective isolation particles of assigned properties from an aerosol stream by selecting the output wavelength within the absorption band of the particle material.

The magnitude of the photophoretic force caused by the collisions of gas molecules with a nonuniformly heated particle surface is generally much greater than the light pressure force. In rare cases it is necessary to consider the combined action of the two forces. In addition, the effect of the reaction of vaporized molecules is significant in some cases.

The theoretical methods used to derive expressions for the photophoretic force and the photophoretic velocity are chosen by comparing the radius $a$ of the particle with the mean free path $\lambda$ of the gas molecules. If the Knudsen number $\mathrm{Kn} = \lambda/a$ is large, then, according to the classification of particles in the physics of disperse systems in air, the particle is termed small. The theory of photophoresis for large Knudsen numbers is devised on the basis of the kinetic theory of gases. The main assumption here is that a particle has little influence on the velocity distribution of the gas surrounding it. The most exact results for the photophoretic force and the photophoretic velocity were presented in Ref. 9. In the expressions for the photophoretic force and the photophoretic velocity of small particles the tangential momentum and energy accommodation coefficients are significant. The results of the experiments in Refs. 8 and 10 are in good agreement with the theory. If the Knudsen number is small, the particle is termed ''coarse,'' and in this case the theory of photophoresis is devised on the basis of the hydrodynamic method, i.e., the hydrodynamic equations and the heat-transfer equations are solved together.[11] At small values of the Reynolds number $\mathrm{Re}=Ua/\nu$, where $U$ is the velocity of the stream of gas flowing past the particle at large distances from it and $\nu$ is the kinematic viscosity, the hydrodynamic equations are replaced by linearized equations.[12,13] The same approach was used to solve the problem of the photophoretic motion of moderately coarse aerosol particles. Detailed reviews of the work on the theory of photophoresis can be found in Refs. 14 and 15. However, in the case of coarse aerosol particles there is some disparity between the theoretical values for the photophoretic velocity and the experimental data. This fact is stimulating a search for new effects, whose inclusion can improve the known models.

The present work examines the photophoretic motion of a coarse solid spherical aerosol particle of nonuniform thermal conductivity suspended in a one-component gas at small Reynolds numbers with consideration of the inertial effects in the hydrodynamic equations (the Oseen method) and the temperature jump in the Knudsen layer.

## EQUATIONS AND BOUNDARY CONDITIONS

We begin the treatment of the motion in a coordinate system, whose origin coincides with the center of gravity of the gas medium. Electromagnetic radiation impinges on the particle and heats its surface. The gas begins to slip along the surface of the particle in the direction of increasing temperature. The thermal slipping gives rise to a photophoretic force. The particle undergoes accelerated motion under the influence of the photophoretic force. When the magnitude of the photophoretic force becomes equal to the magnitude of the force of viscous drag of the medium, the particle begins to move linearly and uniformly with a certain photophoretic velocity $\mathbf{U}_{\mathrm{ph}}$. Because of the small value of the thermal relaxation time, we assume that the heat-transfer process in the particle/gas-medium system is quasistationary. We shall work all the time below in a $0xyz$ coordinate system, whose origin coincides with the center of the particle and whose $0x$ axis is parallel to the propagation direction of the uniform radiation flux impinging on the particle. Going over to spherical coordinates, we shall measure the angle $\Theta$ from the positive direction of the $0x$ axis. In the $0xyz$ coordinate system the particle is stationary, and the gas flows past the particle. The velocity of the gas at infinity $\mathbf{V}_\infty$ is clearly equal to $\mathbf{U}_{\mathrm{ph}}$ with the opposite sign: $\mathbf{V}_\infty=-\mathbf{U}_{\mathrm{ph}}=U\mathbf{i}$. Considering the stationary motion of a one-component gas of nonuniform temperature relative to the particle in the absence of external forces, we arrive at the following equations of motion:[13,14]

$$U\,\frac{\partial \mathbf{V}}{\partial x}=-\frac{1}{\rho}\nabla p+\nu\nabla^2\mathbf{V},\quad \mathrm{div}\mathbf{V}=0, \tag{1}$$

$$V_x|_{r\to\infty}=U,\quad V_y|_{r\to\infty}=0,\quad V_z|_{r\to\infty}=0,\quad p|_{r\to\infty}=p_{e\infty},$$

$$V_r|_{r=a}=0,\quad V_\Theta|_{r=a}=K_{TSl}\frac{\eta_e}{\rho_e T_{e0}a}\left.\left(\frac{\partial T_e}{\partial\Theta}\right)\right|_{r=a}, \tag{2}$$

where $\mathbf{V}$ is the flow velocity, $\rho$ is the density, $p$ is the pressure in the gas, $K_{TSl}$ is the thermal slipping coefficient,[16] $\eta_e$ is the viscosity of the gas medium, $T_e$ is the temperature of the gas (everywhere below the subscript $e$ denotes the gas medium, and $i$ denotes the particle), and $T_{e0}$ is the value of $T_e$ at $r=0$.

Equations (1) were obtained with consideration of the leading inertial terms in the Navier–Stokes equation (the Oseen method[13]). We assume that $V_r$, $V_\Theta$, $p$, and $T_e$ are functions of only $r$ and $\Theta$. In our formulation of the problem $V_\varphi=0$.

The electromagnetic radiation impinging on the particle is absorbed by it and distributed throughout its volume. As a result, sources of thermal energy appear within the particle with a certain density $q_i$, which we also assume to be a function of $r$ and $\Theta$. Therefore, the thermal part of the problem has the following form:[15]

$$\nabla^2 T_e=0, \tag{3}$$

$$\mathrm{div}(\varkappa_i\nabla T_i)+q_i=0, \tag{4}$$

$$\left.\left(\varkappa_i\frac{\partial T_i}{\partial r}\right)\right|_{r=a}=\varkappa_e\left.\frac{\partial T_e}{\partial r}\right|_{r=a},$$

$$(T_e-T_i)|_{r=a}=C_T\lambda\left.\frac{\partial T_e}{\partial r}\right|_{r=a}, \tag{5}$$

$$T_i|_{r\to0}\neq\infty,\quad T_e|_{r\to\infty}=T_{e\infty}. \tag{6}$$

Here $T_i$ is the temperature within the particle; $\varkappa_e$ and $\varkappa_i$ are the thermal conductivities of the gas and the particle, respectively; and $C_T$ is the temperature jump coefficient.[14] We assume that $T_i$ is a function of $r$ and $\Theta$. We assume that the convergence under the conditions at infinity is uniform with respect to $\Theta$. The value of $\varkappa_e$ is assumed to be constant, and $\varkappa_i$ will be discussed below.

## PHOTOPHORESIS OF AN AEROSOL PARTICLE OF UNIFORM THERMAL CONDUCTIVITY

Let us first consider the thermal part of this problem. In this case Eq. (4) of systems (3)–(6) can be written in the form

$$\nabla^2 T_i=-q_i/\varkappa_i. \tag{7}$$

Using the theory of harmonic functions (including the theory of functions that are regular at infinity[17]), we can show on the basis of the maximum principle[17] and Zaremba's principle[18] that this problem has a unique solution. We seek the solution of this problem in a class of expansions of the form

$$T_i=\sum_{n=0}^{\infty} T_{in}(r)P_n(\cos\Theta),$$

$$T_e=\sum_{n=0}^{\infty} T_{en}(r)P_n(\cos\Theta),$$

where the $T_{in}(r)$ and $T_{en}(r)$ are unknown functions, and the $P_n$ are Legendre polynomials.

For the existence of a solution we also assume that

$$q_i=\sum_{n=0}^{\infty} q_{in}(r)P_n(\cos\Theta),$$

where the $q_{in}(r)$ are certain functions, which can be found in terms of $q_i$ on the basis of the orthogonality of the Legendre polynomials according to the following formula:

$$q_{in}(r)=\frac{2n+1}{2}\int_0^{\pi} q_i(r,\Theta)P_n(\cos\theta)\sin\Theta\, d\Theta.$$

We proceed to a proof of the existence of a solution. From Eq. (3) with consideration of (6) we obtain

$$T_e=T_{e\infty}+\sum_{n=0}^{\infty}\frac{A_{en}}{r^{n+1}}P_n(\cos\Theta), \tag{8}$$

where the $A_{en}$ are undetermined coefficients.

From Eq. (7) we obtain the following differential equation for $T_{in}$:

$$r^2T''_{in}+2rT'_{in}-n(n+1)T_{in}=-r^2\frac{q_{in}}{\varkappa_i}. \tag{9}$$

Solving Eq. (9) by standard methods,[19] we obtain

$$T_{in}=\frac{A_{in}}{r^{n+1}}+r^nB_{in}+\frac{1}{(2n+1)\varkappa_i}\times\left[\frac{1}{r^{n+1}}\int_a^r q_{in}r^{n+2}dr-r^n\int_a^r q_{in}\frac{dr}{r^{n-1}}\right].$$

Now substituting the expansions found for $T_i$ and $T_e$ into the boundary conditions (5) and taking into account condition (6), we obtain a combined system of linear equations for determining the coefficients $A_{en}$, $A_{in}$, and $B_{in}$. The existence of a solution has been proved.

We move on to an analysis of the hydrodynamic part [Eqs. (1) and (2)] of the problem. We are interested in the expression for the photophoretic velocity, which, in turn, is derived from the expression for the force. If, instead of Eqs. (1), we consider the equations

$$0=-\frac{1}{\rho}\nabla p+\nu\nabla^2\mathbf{V}, \quad \text{div}\,\mathbf{V}=0, \tag{10}$$

we obtain a generalization of the Stokes method to the nonisothermal case. The boundary condition for thermal slipping (2) now breaks up into an infinite number of equations, since the radial component $V_r$ of the mass velocity $\mathbf{V}$ is sought in the form of the expansion

$$V_r=\sum_{n=1}^{\infty} V_{rn}(r)P_n(\cos\Theta), \tag{11}$$

and the tangential component $V_\Theta$ is sought in the form of the expansion

$$V_\Theta=-\frac{2}{\sin\Theta}\sum_{n=1}^{\infty} V_{\Theta n}(r)J_{n+1}(\cos\Theta), \tag{12}$$

where the $V_{rn}$ and $V_{\Theta n}$ are unknown functions, and the $J_{n+1}$ are Gegenbauer functions.

It is known[11] that to determine the magnitude of the force it is sufficient to determine the first terms (i.e., the terms corresponding to $n=1$) in expansions (11) and (12). Let $V_r^{(1)}$, $V_\Theta^{(1)}$, and $T_e^{(1)}$ be the first terms in (11), (12), and (8), respectively. Then it follows from condition (2) that

$$V_\Theta^{(1)}|_{r=a}=K_{TSl}\frac{\eta_e}{\rho_eT_{e0}a}\left(\frac{\partial T_e^{(1)}}{\partial\Theta}\right)\Bigg|_{r=a}. \tag{13}$$

If we introduce the notation

$$\varepsilon=K_{TSl}\frac{\eta_e}{\rho_eT_{e0}a}\left(\frac{\partial T_e^{(1)}}{\partial\Theta}\right)\Bigg|_{\substack{r=a\\ \Theta=\pi/2}}, \tag{14}$$

the Stokes formula[13] $\mathbf{F}_c=6\pi\eta U_a$ for the force exerted on a spherical particle is generalized in the following manner:

$$\mathbf{F}_c=6\pi\eta U_a\left(1+\frac{2}{3}\frac{\varepsilon}{U}\right)\mathbf{i}. \tag{15}$$

The following formula for the photophoretic velocity can easily be derived from Eq. (15):

$$\mathbf{U}_{\text{ph}}=\frac{2}{3}\varepsilon\mathbf{i}. \tag{16}$$

Thus, from the thermal part of the problem only $T_e^{(1)}$ appears in the expression for the photophoretic velocity. If $\mathbf{V}$ is sought with consideration of the inertial effects in the hydrodynamic equations on the basis of Eqs. (1) and if $T_e$ in condition (2) is replaced by $T_e^{(1)}$, Eq. (15) can be generalized in the following manner:[20]

$$\mathbf{F}_c=6\pi\eta Ua\left(1+\frac{3}{8}\,\text{Re}\right)\left(1+\frac{2}{3}\frac{\varepsilon}{U}\right)\mathbf{i}. \tag{17}$$

It follows from (17) that the photophoretic velocity is insensitive to consideration of the inertial effects in the hydrodynamic equations within the approach described here. It follows from (8) that

$$\frac{\partial T_e}{\partial\Theta}\Bigg|_{\substack{r=a\\ \Theta=\pi/2}}=-\frac{A_{e1}}{a^2}. \tag{18}$$

It was stated during the proof of the existence of a solution that the system of linear equations for determining the coefficients $A_{en}$, $A_{in}$, and $B_{in}$ is a combined system. From this system we can find, in particular, that

$$A_{e1}=\frac{3}{4\pi}\frac{(\mathbf{D},\mathbf{i})}{2\varkappa_e+(1+2C_T\lambda/a)\varkappa_i}, \tag{19}$$

where

$$\mathbf{D}=\left(\int_V(q_i,\mathbf{r},\mathbf{i})dV\right)\mathbf{i} \tag{20}$$

is the dipole moment of the heat-source density (the integration is carried out over the entire volume of the particle $V$).

On the basis of Eqs. (16), (14), and (18)–(20) we obtain

$$\mathbf{U}_{\text{ph}}=\frac{-2\nu_e K_{TSl}}{3T_{e0}(2\varkappa_e+\varkappa_i\beta)}\left(\frac{1}{V}\int_V (q_i,\mathbf{r},\mathbf{i})dV\right)\mathbf{i},$$

$$\beta=1+2C_T\lambda/a. \tag{21}$$

### VARIABLE THERMAL CONDUCTIVITY

Generally speaking, $\varkappa_i$ is a variable which depends on $r$ and $\Theta$. For most natural and artificially produced aerosol particles the dependence of $\varkappa_i$ on $r$ is significant, and the dependence on $\Theta$ is weak, as, for example, when an aerosol particle forms on a condensation nucleus and then solidifies while maintaining its inhomogeneous internal structure, the variation of the thermal conductivity being due to the different compositions of the core and shell of the particle. The dependence of $\varkappa_i$ on $\Theta$ in such particles can be caused by the known dependence of $\varkappa_i$ on $T_i$, which, in turn, is a function of $r$ and $\Theta$. However, the dependence of $\varkappa_i$ on $T_i$ can be neglected. In fact, the real temperature drop $\delta T_i$ over the radius of the particle is of the order of $\delta T_i=a|(\nabla T_e)_\infty|$. Taking into account that $a\sim 10^{-6}$ m, we note that for $\delta T_i$ to be at least of the order of 100 K, the value of $|(\nabla T_e)_\infty|$ must be of the order of $10^8$ K/m, which is never encountered in real aerosol systems on earth. In our opinion $\delta T_i\sim 100$ K is the minimum temperature drop over the radius of a particle, at which the influence of the variation of the temperature on the thermal conductivity of known substances begins to be manifested at all. We, therefore, assume that $\varkappa_i=\varkappa_i(r)$. We construe $\varkappa_i(r)$ as a positive function assigned in a certain half interval $[0, b)$ containing the segment $[0, a)$. We also assume that $\varkappa_i(r)$ can be represented in the form

$$\varkappa_i(r)=\sum_{s=0}^{\infty}\varkappa_{i,s}r^s,\quad r\in[0,\,b). \tag{22}$$

We also assume that the continuation $\varkappa_i^*(t)$ of $\varkappa_i(r)$ onto the complex plane $\mathbb{C}$ on the basis of (22) does not have zeros within the circle $|t|<b$.

### PHOTOPHORESIS OF AN AEROSOL PARTICLE WITH VARIABLE THERMAL CONDUCTIVITY

We start out from the thermal part [Eqs. (3)–(6)] of the problem. In this case Eq. (4) takes on the form

$$\nabla^2 T_i=-\frac{\varkappa_i'}{\varkappa_i}\frac{\partial T_i}{\partial r}-\frac{q_i}{\varkappa_i}. \tag{23}$$

We also seek the solution of this problem in the form of expansions in Legendre polynomials. It is clear that $T_e$ has form (8). To find $T_{in}$, instead of Eq. (9) we obtain the following differential equation:

$$r^2T_{in}''+\left(2+r\frac{\varkappa_i'}{\varkappa_i}\right)rT_{in}'-n(n+1)T_{in}=-r^2\frac{q_{in}}{\varkappa_i}. \tag{24}$$

It follows from (22) that one particular solution $M_{1n}$ of the homogeneous equation corresponding to the inhomogeneous differential equation (24) can be found in a class of generalized power series.[19,21] Let

$$b_0=2,\quad b_1=\frac{\varkappa_{i,1}}{\varkappa_{i,0}},\quad b_s=\frac{s\varkappa_{i,s}-\sum_{j=1}^{s-1}\varkappa_{i,j}b_{s-j}}{\varkappa_{i,0}},$$

where $s\geqslant 2$.

We use the coefficients $b_s$ to define the coefficients $\alpha_s^{(n)}$:

$$\alpha_0^{(n)}=1,\quad \alpha_s^{(n)}=-\frac{\sum_{j=1}^{s}(n+s-j)\alpha_{s-j}^{(n)}b_j}{s(s+2n+1)},$$

where $s\geqslant 1$.

Then it can be shown[19] that the power series

$$M_{1n}=r^n\sum_{s=0}^{\infty}\alpha_s^{(n)}r^s$$

converges in $(0, b)$ and satisfies Eq. (24). The second solution $M_{2n}$ (of the corresponding homogeneous equation), which is linearly independent with $M_{1n}$, can be found using the formula[19]

$$M_{2n}=M_{1n}\int_a^r\frac{dr}{r^2\varkappa_i M_{1n}^2}.$$

We note that $M_{1n}\sim r^n$ when $r\rightarrow 0$, and if $\varkappa_i\equiv\text{const}$, then $M_{1n}=r^n$. The functions $M_{1n}$ and $M_{2n}$ form a fundamental system of equations of the homogeneous equation corresponding to Eq. (24). The general solution of Eq. (24) can be written in terms of these functions in the form[19]

$$T_{in}=A_{in}M_{2n}+B_{in}M_{1n}-M_{1n}\times\int_a^r M_{2n}q_{in}r^2dr+M_{2n}\int_a^r M_{1n}q_{in}r^2dr.$$

Now substituting $T_e$ and $T_i$ into the boundary conditions (5) and taking into account the boundary condition for $T_i$ from (6), we obtain the combined system of linear equations for determining the coefficients $A_{en}$, $A_{in}$, and $B_{in}$. Thus, the existence of a solution has been proved. In particular,

$$A_{e1}=\frac{3}{4\pi}\frac{(\mathbf{D}',\mathbf{i})}{T_{e0}[2\varkappa_e+\gamma\varkappa_i(a)\beta]}, \tag{25}$$

where

$$\mathbf{D}'=\left(\int_V(m(r)q_i\mathbf{r},\mathbf{i})dV\right)\mathbf{i},$$

$$m(r)=\frac{aM_{11}(r)}{rM_{11}(a)},\quad \gamma=a\frac{M_{11}'(a)}{M_{11}(a)}.$$

We find the photophoretic velocity on the basis of Eqs. (16), (14), (18), and (25)

$$\mathbf{U}_{\text{ph}}=\frac{-2\nu_e K_{TSl}}{3T_{e0}[2\varkappa_e+\gamma\varkappa_i(a)\beta]}\left(\frac{1}{V}\int_V(m(r)q_i,\mathbf{r},\mathbf{i})dV\right)\mathbf{i}. \tag{26}$$

Let $\alpha=\varkappa_e/\varkappa_i(a)$. Then $\mathbf{U}_{\rm ph}$ can be expressed in terms of $\mathbf{U}_{\rm ph}|_{\varkappa_i\equiv\varkappa_i(a)}$, which is calculated from Eq. (21) with $\varkappa_i\equiv\varkappa_i(a)$, in the following manner:

$$\mathbf{U}_{\rm ph}=\mathbf{U}_{\rm ph}|_{\varkappa_i\equiv\varkappa_i(a)}\left\{1+\frac{(R-1)(\beta+2\alpha)+(1-\gamma)\beta}{2\alpha+\beta\gamma}\right\}, \tag{27}$$

where

$$R=\int_V (m(r)q_i\mathbf{r},\mathbf{i})dV\bigg/\int_V (q_i\mathbf{r},\mathbf{i})dV.$$

We note that if $\varkappa_i\equiv\varkappa_i(a)$, $M_{11}(r)\equiv r$; therefore, $\gamma=1$, $R=1$, and in this case Eq. (26) coincides with Eq. (21).

## BLACKBODY

When the particle absorbs electromagnetic radiation as a blackbody, absorption occurs in a thin layer of thickness $\delta\ll a$ adjacent to the heated part of the surface. If $E$ is the intensity of the incident radiation, then

$$q_i(r,\theta)=\begin{cases} 0, & 0\leqslant\Theta<\dfrac{\pi}{2}, \quad 0\leqslant r\leqslant a,\\[2ex] 0, & \dfrac{\pi}{2}\leqslant\Theta\leqslant\pi, \quad 0\leqslant r<a-\delta,\\[2ex] -\dfrac{E}{\delta}\cos\Theta, & \dfrac{\pi}{2}\leqslant\Theta\leqslant\pi, \quad a-\delta\leqslant r\leqslant a. \end{cases}$$

Therefore, in the case of a blackbody, Eq. (26) takes on the following form:

$$\mathbf{U}_{\rm ph}=K_{TSl}\frac{\nu_e}{3T_{e0}[2\varkappa_e+\gamma\varkappa_i(a)\beta]}E\mathbf{i}.$$

For a blackbody Eq. (27) is simplified to

$$\mathbf{U}_{\rm ph}=\mathbf{U}_{\rm ph}|_{\varkappa_i\equiv\varkappa_i(a)}(1+f(\alpha,\beta,\gamma)),$$

$$f(\alpha,\beta,\gamma)=\frac{(1-\gamma)\beta}{2\alpha+\beta\gamma}.$$

Let $10^{-3}\leqslant\alpha\leqslant1$, $1\leqslant\beta\leqslant1.6$, and $0.2\leqslant\gamma\leqslant8$. The condition on $\alpha$ covers a broad spectrum of medium-particle combinations. The condition on $\beta$ covers the region of small Knudsen numbers. We shall impart a physical meaning to the condition on $\gamma$ in the next section. Investigating the function $f(\alpha,\beta,\gamma)$ under the conditions on $\alpha$, $\beta$, and $\gamma$ indicated above, we find that $f(\alpha,\beta,\gamma)$ varies in the range from $-0.8$ to 4. Thus, the variable internal thermal conductivity $\varkappa_i=\varkappa_i(r)$ can introduce significant changes into the photophoretic velocity calculated for $\varkappa_i\equiv\varkappa_i(a)$.

## MODEL PARTICLE WITH PRONOUNCED INHOMOGENEITY

As a model of such a particle we consider the dependence

$$\varkappa_i(r)=\varkappa_i(0)\exp(kr),$$

where $k$ is an arbitrary constant.

In this case it can be shown that

$$\gamma=\left(-2+x\,\frac{\exp(x)-1-x}{\exp(x)-1-x-0.5x^2}\right)\bigg|_{x=-ak},$$

$$\frac{M_{11}(r)}{r}=6\left(\frac{\exp(x)-1-x-0.5x^2}{x^3}\right)\bigg|_{x=-kr}.$$

Thus, here $\gamma$ and $M_{11}(r)$ have analytical representations. In the case of a blackbody, whose thermal conductivity obeys an exponential law, it is easy to show that the inequality $0.2\leqslant\gamma\leqslant8$ is equivalent to the inequality $-10\leqslant ak\leqslant10$. However, $ak$ characterizes the relative drop in the thermal conductivity over a distance equal to one particle radius. Therefore, we now have a physical interpretation for the condition on $\gamma$ from the preceding section. This condition covers a broad spectrum of relative drops in the thermal conductivity over a distance equal to one particle radius.

## PRINCIPAL CONCLUSIONS

The velocity of the particle remains insensitive to consideration of the inertial effects in the hydrodynamic equations. Since the influence of the variable internal thermal conductivity $\varkappa_i=\varkappa_i(r)$ of a particle on its photophoretic motion can be significant, only homogeneous aerosol particles obtained by artificial means should be used in experiments. Particles of nonuniform thermal conductivity that can be described by the model that we selected with an internal thermal conductivity $\varkappa_i=\varkappa_i(r)$ and a volume density of heat sources within the particle $q_i(r,\Theta)$ can be regarded as homogenous particles with a constant value of the thermal conductivity $\gamma\varkappa_i(a)$ and a volume density of heat sources within the particle $q_i(r,\Theta)aM_{11}(r)/[rM_{11}(a)]$.

[1] F. Ehrenhaft, Ann. Phys. **55**, 81 (1918).

[2] V. B. Kutukov and Yu. I. Yalamov, in *Nonlinear Effects in the Propagation of Laser Radiation in the Atmosphere* [in Russian], Tomsk (1977), pp. 145–147.

[3] O. Freining, in *Aerosol Science, No. 4*, edited by C. N. Davies, Academic Press, New York (1966).

[4] M. G. Markov, Candidate's dissertation, Moscow (1983), 179 pp.

[5] G. M. Higy and J. R. Brock, J. Geophys. Res. **72**, 455 (1967).

[6] S. P. Lin, J. Colloid Interface Sci. **51**, 66 (1975).

[7] E. J. Keng and C. G. Orr, Nature **200**, 352 (1963).

[8] N. T. Tong, J. Colloid Interface Sci. **51**, 143 (1975).

[9] V. B. Kutukov, E. R. Shchukin, and Yu. I. Yalamov, Zh. Tekh. Fiz. **46**, 626 (1976) [Sov. Phys. Tech. Phys. **21**, 361 (1976)].

[10] A. I. Bogoletov and G. G. Bystryĭ, in *Reports to the 14th All-Union Conference on Current Topics in the Physics of Aerosol Systems* [in Russian], Odessa University, Odessa (1986), pp. 153–154.

[11] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics*, Noordhoff, Leyden (1973) [Russian trans. Mir, Moscow (1976), 632 pp.].

[12] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., Pergamon Press, Oxford (1987) [Russian original, Nauka, Moscow (1988), 736 pp.].

[13] N. E. Kochin, I. A. Kibel', and N. V. Roze, *Theoretical Hydromechanics*, Interscience Publishers, New York (1964) [Russian original, OGIZ, Moscow (1948), 612 pp.].

[14] Yu. I. Yalamov and V. S. Galoyan, *Dynamics of Droplets in Inhomogeneous Viscous Media* [in Russian], Luĭs, Moscow (1985), 205 pp.

[15] E. R. Shchukin, Yu. I. Yalamov, and Z. L. Shulimanova, *Selected Topics in Aerosol Physics* [in Russian], MPU, Moscow (1992), 297 pp.

[16] I. I. Ivchenko and Yu. I. Yalamov, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 6, 59 (1969).

[17] A. N. Tikhonov and A. A. Samarskiĭ, *Equations of Mathematical Physics*, Pergamon Press, Oxford (1964), Nauka, Moscow, 1972, 735 pp.
[18] A. V. Bitsadze and D. F. Kalinichenko, *A Collection of Problems on the Equations of Mathematical Physics*, Mir, Moscow (1977), 222 pp.
[19] V. I. Smirnov, *A Course of Higher Mathematics, Vol. 2: Advanced Calculus*, Addison-Wesley, Reading, Mass. (1964) [Russian original, Nauka, Moscow (1967), 655 pp.].
[20] Yu. I. Yalamov and A. S. Khasanov, ‘‘Thermophoresis of a solid spherical coarse aerosol particle with consideration of the inertial effects in the hydrodynamic equations’’ [in Russian], Document deposited in the All-Union Institute of Scientific and Technical Information of the Russian Academy of Sciences (VINITI), No. 3196-V95, 1995.
[21] M. A. Lavrent’ev and B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable* [in Russian], Nauka, Moscow (1973), 736 pp.

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Lagrange formalism for particles moving in a space of fractal dimension

I. P. Guk

*Institute of Pulse Processes and Technologies, Ukrainian National Academy of Sciences, 327018 Nikolaev, Ukraine*

(Submitted November 11, 1996)
Zh. Tekh. Fiz. **68**, 7–11 (February 1990)

Analogs of the Lagrange equation for particles evolving in a space of fractal dimension are obtained. Two cases are considered: 1) when the space is formed by a set of material points (a so-called fractal continuum), and 2) when the space is a true fractal. In the latter case the fractional integrodifferential formalism is utilized, and a new principle for devising a fractal theory, viz., a generalized principle of least action, is proposed and used to obtain the corresponding Lagrange equation. The Lagrangians for a free particle and a closed system of interacting particles moving in a fractal continuum are derived. © *1998 American Institute of Physics.* [S1063-7842(98)00204-9]

## INTRODUCTION

Such physical processes as the explosion of wires and the breakdown of insulators belong to a single class of phenomena, which are known as critical phenomena. Such processes are described using the methods of the theory of critical phenomena, renormalization group analysis, and series expansions according to approximation theory.[1,2] One of the most important postulates of the theory of critical phenomena is scale invariance, which essentially involves the introduction of a new symmetry, viz., similarity symmetry, into a space. This postulate forms the basis of renormalization group analysis. However, it is now more important for us that when similarity (more precisely, self-similarity) properties are introduced in a space, it can be classified as an object with a fractal structure. Therefore, the description of the explosion of wires and the breakdown of insulators within the concept of fractals would be of interest.

Theoretical investigations of various physical phenomena on fractals (more precisely, on systems of the fractal type) and within fractals have been based on the fractional integrodifferential formalism,[3–5] nonstationary analysis,[6] etc. This has been motivated by the following statement:[4] the propagation of particles and waves in true fractal media should be described by other, more general equations, which go over to the conventional linear equations for smooth media.

The diverse investigations of fractals (as a conception) have reached a new level, i.e., the systemization and ordering of the information that has been accumulated. The methods and possibilities of the new theories are gradually taking shape.[4,5,7] However, the lack of fundamental principles has delayed the formulation of a general fractal theory. Another delaying factor is the absence of generally accepted mathematical machinery for treating problems involving fractals. The use of the fractional integrodifferential formalism was proposed for this purpose in Refs. 4 and 8.

In this paper the equation of motion of a particle in a fractal continuum (the definition is given in Sec. 1) is obtained on the basis of the principle of least action. If a space (or a spatial object) is a true fractal, the equation of motion can be obtained on the basis of the generalized principle of least action, which is proposed as a fundamental principle for devising a fractal theory. The concept of an inertial fractal reference frame, analogs of Galilean transformations, and the Galilean relativity principle are introduced in a space of fractal dimension. The Lagrangian for a free particle and a closed system of interacting particles moving in the fractal continuum is also derived.

## 1. MODIFIED LAGRANGE FORMALISM

Let there be a set of material points comprising a space, in which an observable point particle can be detected as it evolves. We impart the property of self-similarity to the set of points introduced, and we call such a set a fractal continuum.

We now formulate the problem as the description of the motion of a certain material point in a fractal continuum. To introduce an analog of action from ordinary mechanics,[9] the following condition for the variable $t$ (the time) must be added: it must ''scan'' all motion phenomena. This requires fulfillment of the following condition: each generalized coordinate must be a continuous function together with the first derivative with respect to the scanning variable. Our next task is to introduce the generalized principle of least action.

## 2. GENERALIZED PRINCIPLE OF LEAST ACTION

Let there be a set of generalized coordinates $\kappa$ and a so-called scanning variable $\tau$, with respect to which the set of coordinates $\kappa$ and their first derivatives are continuous. In other words, the evolution of the generalized coordinates is observed in terms of $\tau$. The rate of variation of $\kappa$ is determined by the ordinary derivative $d\kappa/d\tau$. We now postulate that the state of the system is described completely by assigning the entire set of generalized coordinates and the generalized velocities $\kappa_\tau$. Then the generalized principle of least action can be formulated by the following statement: each mechanical system is characterized by a definite function $L(\kappa,\kappa_\tau,\tau)$, which is continuous, as is the first derivative, and by $\tau$. Motion of the system occurs under the following condition.

Let the system occupy positions characterized by the two sets of values of the generalized coordinates $\kappa_{\{1\}}$ and $\kappa_{\{2\}}$ at the values of the scanning variable $\tau=\tau_1$ and $\tau=\tau_2$. Then, the system moves between these positions in such a manner that the integral

$$S=\int_{\tau_1}^{\tau_2} L(\kappa,\dot{\kappa}_\tau,t)\,d\tau \tag{1}$$

would have the smallest possible value (more precisely, an extremum value). The function $L$ is called a generalized Lagrangian, and $S$ is called a generalized action.

## 3. LAGRANGE EQUATION IN TERMS OF THE FRACTIONAL INTEGRODIFFERENTIAL FORMALISM

We consider the motion of a certain system as evolution in terms of the scanning variable $\tau$ (when $\tau$ is the time, we refer to temporal evolution, but in the general case $\tau$ is not identified with the time). We vary the action according to a standard scheme (see, for example, Ref. 10):

$$\delta S=\int_{\tau_1}^{\tau_2} L(\kappa+\delta\kappa,\kappa_\tau+\delta\kappa_\tau,\tau)\,d\tau - \int^{\tau_2}\tau_1 L(\kappa,\kappa_\tau,\tau)\,d\tau=0. \tag{2}$$

As a result, we obtain the Lagrange-Euler equation

$$\frac{\partial L}{\partial\kappa}-\frac{d}{d\tau}\frac{\partial L}{\partial\kappa_\tau}=0. \tag{3}$$

At this point we note that in the case of a fractal continuum the relationship between the generalized coordinates $\kappa_l$ and the coordinates of an ordinary space $x_l$ has a scaling character and that the scanning variable $\tau$ is the time $t$ (in the simplest case of the type $\kappa_l \sim x_l^D$, where $l$ labels the coordinates and $d$ is the fractal dimension). Therefore, when we go over to the laboratory reference frame of an ordinary space, we obtain equations of the type (such an equation is an analog of the Lagrange equation in a fractal continuum)

$$\frac{\partial L}{\partial x_l^D}-\frac{d}{dt}\left(\frac{\partial L}{\partial x_{\tau_l}^D}\right)=0. \tag{4}$$

We perform the limiting transition from a fractal continuum to a true fractal by going from the smooth function $L(\kappa,\kappa_\tau,\tau)$ to the fractal function $L(x,\dot{x},t)$, which is everywhere nondifferentiable. The transition is accomplished by replacing the ordinary derivatives with Riemann–Liouville derivatives of fractional order

$$\hat{D}_x^D[L(x,\dot{x},t)]-\frac{d}{dt}\hat{D}_{\dot{x}}^D[L(x,\dot{x},t)]=0. \tag{5}$$

Expression (5) is an analog of the Lagrange equation in terms of the fractional integrodifferential formalism.

## 4. PRINCIPLE OF RELATIVITY IN A FRACTAL CONTINUUM

Let us assume that for a fractal continuum to be homogeneous and isotropic relative to an ‘‘inertial’’ reference frame, the reference frame itself must be a fractal of the same dimension as the fractal continuum. To fix ideas, we call such a reference frame an inertial fractal reference frame, which is inertial in the sense that the fractal continuum is homogeneous and isotropic relative to it, and the scanning variable $\tau$ is homogeneous.

The form of the Lagrangian $L(\kappa,\dot{\kappa}_\tau,t)$ of a free material point in the inertial fractal reference frame is specified by the following arguments: the homogeneity of the fractal continuum and the scanning variable means that $L$ does not contain explicit dependence on $\kappa$ and $\tau$, and the isotropy indicates that there is no dependence on the direction of the generalized velocity vector (in other words, $L$ depends on a scalar quantity formed by components of the generalized velocity, e.g., $\Sigma_l\kappa_\tau^l\kappa_{\tau_l}$), i.e., $L=L(|\kappa_\tau|^2)$. In this case the Lagrange equations take on the form

$$\frac{d}{d\tau}\frac{\partial L}{\partial\kappa_\tau^l}=0\to\frac{\partial L}{\partial\kappa_\tau^l}=\text{const}, \tag{6}$$

where the superscript $l$ labels the components of the generalized coordinates (or velocities).

It is taken into account here that $\partial L/\partial\kappa^l=0$ under the assumption of homogeneity of the fractal continuum. Equation (6) has the following solution

$$\kappa_\tau^l=\text{const}. \tag{7}$$

Thus, in an inertial fractal reference frame any free material point moves so that the components of the generalized velocity are conserved. This result is a generalization of the familiar law of inertia in classical mechanics.

Let us now consider two different initial fractal reference frames: K and K′, the latter of which moves with a constant generalized velocity $\{X_\tau^l\}$. A material point has the sets of generalized coordinates $\{\kappa_l\}$ and $\{\kappa_l'\}$ in the K and K′ reference frames. There is a relationship between the sets of coordinates (the evolution of the scanning variable is assumed to be identical in both inertial fractal reference frames):

$$\kappa^l=\kappa'^l+X_\tau^l\tau,\qquad \tau=\tau'. \tag{8}$$

The transformations (8) of the generalized coordinates are analogous to the familiar Galilean transformations. The requirement that the equations of motion of mechanics in a fractal continuum be invariant with respect to the transformations (8) is a generalization of the Galilean relativity principle.

## 5. LAGRANGIAN OF A FREE MATERIAL POINT

In developing a mechanics in a space with a noninteger dimension we shall bear in mind that upon asymptotic passage to an ordinary (one-, two-, or three-dimensional) space the form of the functions and the equations describing the motion of a material point must coincide with the classical expressions.

We proceed to constructing the Lagrangian of a free material point moving in a fractal continuum. The reasoning is analogous to that in Ref. 8. The function $L$ of a free material point does not depend on the scanning variable $\tau$, the generalized coordinates $\kappa$, or the direction of the generalized velocity vector. Thus, $L$ depends on the square of the absolute value of the generalized velocity:

$$L \sim |\kappa_\tau|^2. \tag{9}$$

The relationship to the laboratory coordinate frame, in which the position of the point is specified by the coordinate $x$, is defined by the expression ($\alpha$ is a proportionality factor)

$$\kappa = \alpha x^D, \quad \kappa_\tau = \alpha x^{D-1} x_\tau. \tag{10}$$

The Lagrangian of a free material point is systematically brought into the form

$$L = \frac{m' \kappa_\tau^2}{2}, \tag{11}$$

$$L = \frac{m}{2}(x^{D-1} x_\tau) = L = \frac{m v^2}{2} x^{2D-2},$$

$$m = m' \alpha^2, \quad v = x_\tau, \tag{12}$$

where $m$ and $m'$ are constants.

Assuming that the fractal dimension differs only slightly from $D=1$, we expand the function into a series in the small parameter $(D-1)$:

$$\begin{aligned} L &= L_0 \left( 1 + \left. \frac{\partial x^{2D-1}}{\partial D} \right|_{D=1} (D-1) \right. \\ &\quad \left. + \left. \frac{\partial^2 x^{2D-1}}{\partial D^2} \right|_{D=1} (D-1)^2 + \ldots \right) \\ &= L_0 (1 + 2(D-1)\ln(x) + \{2(D-1)\ln(x)\}^2 + \ldots) \\ &= L_0 \left( \sum_{n=0}^{\infty} \{2(D-1)\ln(x)\}^n \right), \end{aligned} \tag{13}$$

where $L_0 = (m v^2)/2$.

If $|2(D-1)\ln(x)| < 1$, the expression takes the following form:

$$L = \frac{L_0}{1 - 2(D-1)\ln(x)}. \tag{14}$$

The result is not, in principle, unexpected, since it is known that a term with $\ln(x)$ appears when quasi-two-dimensional (quasi-one-dimensional) structures are considered.[11]

## 6. LAGRANGIAN OF A SYSTEM OF MATERIAL POINTS

An expression for the Lagrangian of a free particle was derived above. It can be obtained for a system of noninteracting particles, if we take into account the additivity of $L$:

$$L = \sum_{b=0}^{N} \frac{m_b' (\kappa_\tau)_b^2}{2}. \tag{15}$$

The interaction of the particles can be taken into account after Ref. 8 by adding a definite function of the generalized coordinates to (15) (as follows from the instantaneous nature of interaction transfer):

$$L = \sum_{b=0}^{N} \frac{m_b' (\kappa_\tau)_b^2}{2} - U(\overline{\kappa_1}, \overline{\kappa_2}, \overline{\kappa_3}, \ldots), \tag{16}$$

where $N$ is the number of particles and $\overline{\kappa_l}$ is the radius vector of a particle in the fractal continuum.

The equations of motion are obtained by plugging (16) into the generalized Lagrange equations:

$$\frac{d}{d\tau} \frac{\partial L}{\partial \kappa_\tau^l} = \frac{\partial L}{\partial \kappa^l} \rightarrow m_b \frac{d \kappa_\tau^l}{d\tau} = -\frac{\partial U}{\partial \kappa^l}, \tag{17}$$

where the superscript $l$ labels the components of the vector.

The transition to the laboratory reference frame is accomplished by transformations of the coordinates:

$$\kappa_l = \sum_{i=0} \beta_{li} x_i^{D_{il}}. \tag{18}$$

Then the equations of motion in the laboratory reference frame can be written in the form

$$\frac{1}{\Sigma_{i=0} \beta_{li} D_{il} \nu_i^{D_{il}-1} w_i} = \sum_{i=0} \beta_{li} D_{il} \kappa_i^{D_{il}-1} \frac{1}{F_i}, \tag{19}$$

where we have introduced the notation for the generalized velocity, acceleration, and force

$$\nu_i = \frac{\partial \kappa_i}{\partial \tau}; \quad w_i = \frac{\partial^2 \kappa_i}{\partial \tau^2}; \quad F_i = \frac{\partial U}{\partial \kappa_i}, \tag{20}$$

i.e., in a fractal continuum the relationship between the force acting on a particle and the acceleration is not the directly proportional relation (19) in the general case.

## CONCLUSIONS

A space which is formed by a set of material points and has a fractal structure cannot be identified with a true fractal; therefore, such an object was termed a fractal continuum. Such objects are observed in media in which phase transitions occur (for example, the explosion of wires and the breakdown of insulators). The equation of motion of particles within such an object has been derived on the basis of a generalized principle of least action. In a theoretical study of the motion of particles (quasiparticles) in a true fractal, devising a Lagrange formalism is complicated by the fact that the function describing such motion belongs to the class of fractal functions. As we know, fractal functions are everywhere nondifferentiable; therefore, the usual procedure for finding the action extremum cannot be used. A way out of this situation has been proposed on the basis of a generalized principle of least action, which contemplates the introduction of a scanning variable, with respect to which the Lagrangian is continuous.[12] For example, we offer the following problem. ‘‘A biological object, viz., a cell with a set of chromosomes, is given. We make a mark on some fragment of one particular chromosome. Two sets of chromosomes form during cell division (in the first stage). Let $N$th divisions take place. Determine the location of the marked fragment.'' The set of points (in this case a point is a cell nucleus) is isomorphic to a Cantor set. The role of the scanning variable can be played by the number of the act of division.

Generalization of the law of inertia, the Galilean relativity principle, and Galilean transformations have been obtained by introducing the concept of an inertial fractal reference frame into a space of fractal dimension.

Expressions for the Lagrangians of a free particle and a system of interacting particles, whose motion takes place in a self-similar space of noninteger dimension, have also been obtained in terms of the laboratory reference frame. In contrast to an ordinary isotropic space of integer dimension, here there is a dependence on the radius vector of the laboratory reference frame and the dimension of the fractal continuum.

## APPENDIX: FRACTIONAL INTEGRODIFFERENTIAL FORMALISM

The mathematical machinery of fractional integration and differentiation is already sufficiently developed for applying it to problems in theoretical physics. However, this machinery has not yet been widely employed.

A fractional Riemann–Liouville integral is defined by the following expression

$$\hat{I}_x^z\Big[(f(x)]=\frac{1}{\Gamma(z)}\int_b^x \frac{f(t)}{(x-t)^{1-x}}dt. \qquad \text{(A1)}$$

Definition (7) is a generalization of the identity

$$\underbrace{\int_b^x dx\ldots\int_b^x dx\int_b^x}_{n} f(x)dx=\frac{1}{(n-1)!}\int_b^x (x-t)^{n-1}f(t)dt. \qquad \text{(A2)}$$

A fractional integral satisfies the identity

$$\hat{I}_x^z\cdot\hat{I}_x^b=\hat{I}_x^b\cdot\hat{I}_x^z=\hat{I}_x^{z+b}. \qquad \text{(A3)}$$

A fractional Riemann–Liouville derivative is introduced in analogy to a fractional integral:

$$\hat{D}_x^z[f(x)]=\frac{1}{\Gamma(z)}\frac{d}{dx}\int_b^x \frac{f(t)}{(x-t)^{1-z}}dt. \qquad \text{(A4)}$$

Fractional derivatives obey the following identity:

$$\hat{D}_x^z\cdot\hat{D}_x^b=\hat{D}_x^b\cdot\hat{D}_x^z=\hat{D}_x^{z+b}. \qquad \text{(A5)}$$

Fractional integrodifferential operators combine according to the following rules:

$$\hat{D}_x^z[\hat{I}_x^z[f(x)]]=f(x),$$

$$\hat{I}_x^z[\hat{D}_x^z[f(x)]]=f(x)-\sum_{k=1}^{[z]}\hat{D}_x^{z-k}[f(x)]\Big|_{x=z}\frac{(x-z)^{z-k}}{\gamma(z-k+1)}. \qquad \text{(A6)}$$

Laplace and fractional integration operators can be combined:

$$\hat{L}[\hat{I}_x^z[f(x)]]=p^{-z}\hat{L}[f(x)]. \qquad \text{(A7)}$$

In conclusion, we introduce a generalized Leibnitz rule

$$\hat{D}_x^z[f(x)g(x)]=\sum_{k=0}^{\infty}\binom{z}{k}\hat{D}_x^{z-k}[f(x)]\hat{D}_x^k[g(x)], \qquad \text{(A8)}$$

$$\hat{D}_x^z[f(x)g(x)]=\sum_{k=-\infty}^{\infty}\binom{z}{k+b}\hat{D}_x^{z-b-k}[f(x)]\hat{D}_x^{b+k}[g(x)], \qquad \text{(A9)}$$

where the generalized binomial coefficient equals

$$\binom{z}{b}=\frac{\Gamma(z+1)}{\Gamma(b+1)\Gamma(z-b+1)}\,\frac{\sin[\pi(b-z)]\Gamma(z+1)\Gamma(b-z)}{\Gamma(b+1)}, \qquad \text{(A10)}$$

and a Leibnitz equation with a residual term

$$\hat{D}_x^z[f(x)g(x)]=\sum_{k=0}^{n-1}\binom{z}{k}\hat{D}_x^{z-k}[f(x)]\hat{D}_x^k[g(x)]+R_n, \qquad \text{(A11)}$$

where

$$R_n=\frac{(-1)^n}{\Gamma(-z)(n-1)!}\int_z^x (x-t)^{-z-1}f(t)dt \times\int_t^x (x-s)^{n-1}\hat{D}_s^n[g(s)]ds.$$

[1] A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions*, Pergamon Press, Oxford (1979) [Russian original, Nauka, Moscow (1982), 382 pp.].

[2] Shang-Keng Ma, *Modern Theory of Critical Phenomena*, W. A. Benjamin, Reading, Mass. (1976) [Russian translation, Mir, Moscow (1980), 298 pp.].

[3] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications*, Gordon and Breach, Chur–Philadelphia (1993) [Russian original, Nauka i Tekhnika, Minsk (1987), 687 pp.].

[4] A. E. Dubinov (ed.), *Fractals in Applied Physics* [in Russian], All-Russia Scientific-Research Institute of Experimental Physics, Arzamas-16 (1995), 216 pp.

[5] A. I. Olemskoĭ and A. Ya. Flat, Usp. Fiz. Nauk **163**(12), 1 (1993) [Phys. Usp. **36**, 1087 (1993)].

[6] L. Notale and I. Schnelder, J. Math. Phys. **25**, 1296 (1994).

[7] B. M. Smirnov, *Physics of Fractal Clusters* [in Russian], Nauka, Moscow (1991), 133 pp.

[8] Ya. B. Zel'dovich and D. D. Sokolov, Usp. Fiz. Nauk **146**, 493 (1985) [Sov. Phys. Usp. **28**, 608 (1985)].

[9] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, Oxford (1976) [Russian original, Nauka, Moscow (1988), 216 pp.].

[10] H. Goldstein, *Classical Mechanics*, Addison-Wesley Press, Cambridge, Mass. (1950) [Russian trans., Nauka, Moscow (1975), 416 pp.].

[11] T. Ando, A. B. Fowler, and F. Stern, ''Electronic properties of two-dimensional systems,'' Rev. Mod. Phys. **54**, 437–672 (1982) [Russian trans., Mir, Moscow (1985), 416 pp.].

[12] I. P. Guk, in *Material Science and Material Properties for Infrared Optoelectronics, 30 September–2 October 1966, Uzhgorod, Ukraine*, SPIE, Bellingham, Washington (1996), p. 78.

Translated by P. Shelnitz

# Processes involving a change in the charge states during interactions of $He^{2+}$ ions with fullerenes

V. V. Afrosimov, A. A. Basalaev, V. P. Belik, Yu. V. Maĭdl', and M. N. Panov

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

O. V. Smirnov

*St. Petersburg State Technical University, 195251 St. Petersburg, Russia*

(Submitted January 15, 1997)
Zh. Tekh. Fiz. **68**, 12–14 (April 1998)

The cross sections for the elementary processes involving a change of the charge states of both particles during the interaction of $He^{2+}$ ions with fullerene molecules are for the first time measured over a broad energy range of electron-volt energies. It is found that processes involving the capture of one or two electrons by the $He^{2+}$ ions are accompanied by additional ionization of the fullerene and that the collisional contribution of the transfer-ionization processes increases with increasing velocity. Single-electron capture is rarely accompanied by fragmentation of the fullerene. Double-electron capture leads, with a higher probability, to fragmentation with the formation of several light charged fragments and, with a smaller probability, to fragmentation with the formation of a heavy charged fragment containing an even number of carbon atoms and light fragments in an uncharged state. © *1998 American Institute of Physics.* [S1063-7842(98)00304-3]

## 1. INTRODUCTION

Active research has been carried out on the interaction of fullerenes and their ions with electrons[1,2] and ions.[3–6] Mass-spectrometric analyses of the ions formed have revealed that $C_{60}^{q+}$ ions with $q=1-5$ form in collisions of fullerenes with both ions and electrons. It was shown in an investigation of the electron-impact ionization of fullerenes[1] that fragmentation is significant only in processes with multiple ionization of the molecule. In collisions with slow (10–100 eV) $Na^+$ and $Ne^+$ ions the role of the processes with fragmentation of the fullerene is relatively minor in comparison with the formation of a singly charged $C_{60}^+$ ion.[4,5]

The precision analysis of the energy of the ions detected after interactions with fullerenes in Ref. 7 showed that the pure excitation and ionization of a fullerene in the kiloelectron-volt collision-energy range are negligible in comparison with the single- and double-electron-capture processes.

All of these investigations were based on an analysis of the charge state of only one of the collision partners, with the exception of Ref. 3, in which measurements were performed with the detection of both particles, but only at a single value of the collision energy. The purpose of the present work was to directly investigate processes with alteration of the charge state of both collision partners upon the capture of electrons by ions over a broad range of collision velocities in the case of fullerenes. These processes can clearly lead to both additional ionization of the target molecule and its fragmentation. In the case of $He^{2+}$ ions the scheme of the processes investigated can be written down as

$$He^{2+}+C_{60}\rightarrow He^{(2-k)+}+C_{60-x}^{(q-l)+}+C_s^{l+}+(q-k)e^-, \quad (1)$$

where $k=1,2$ is the number of electrons captured by the projectile; $(q-k)$ is the number of free electrons formed as a result of electron capture with ionization of the target (transfer ionization); $C_s^{l+}$ is a fragment containing $s$ carbon atoms that was formed upon fragmentation of the molecule; and $l=0,1,\ldots$ is the charge of the fragment.

The generally accepted notation containing the initial and final charge states of both collision partners $\{20(2-k)q\}$ is used below to denote the elementary processes without fragmentation.

## 2. EXPERIMENTAL METHOD

A monokinetic, well collimated beam of $^3He^{2+}$ ions crossed a jet of fullerene molecules emerging from the capillary of an effusion source heated to 450–480 °C. The charge state of the ions in the beam after the collisions was analyzed using a 30° cylindrical electrostatic analyzer. The slow ions were drawn out of the interaction zone by a uniform electric field with an intensity of 50 V/cm and were directed into a time-of-flight analyzer. The potentials of the ion-optical system of the analyzer were adjusted so as to minimize the spread of the times of flight of the ions formed in the interaction zone of the primary beam with the fullerene jet.

The cross sections for specific elementary processes of type (1) were determined from the experimentally measured spectra of the delay time between detection of the projectile and the recoil particle participating in the same collision. The detection scheme included provisions for the possibility of introducing a regulated time delay between the signals in each of the two particle detection channels. This permitted separation of the processes accompanied by fragmentation

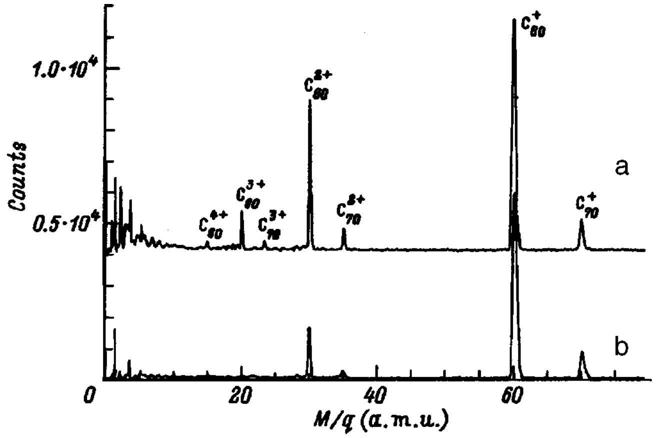

FIG. 1. Mass spectrum of the recoil ions formed in single-electron-capture processes. Energy of the $^3He^{2+}$ ions: $E=68$ (a), 6 keV (b).

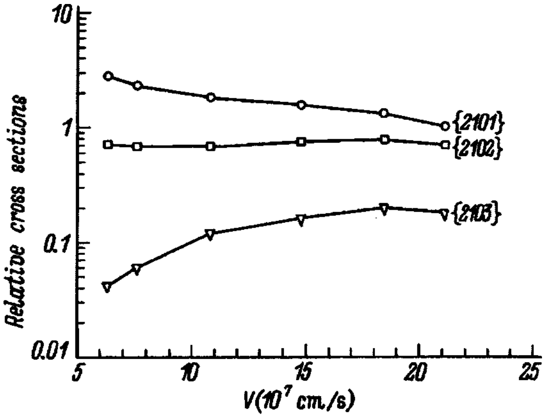

FIG. 2. Relative cross sections for elementary $\{201q\}$ processes due to single-electron capture: ○—{2011}, □—{2012}, ▽—{2013}.

into one charged fragment and one neutral fragment from the processes with fragmentation into two or more charged fragments of different mass. All the data presented below on the cross sections are given in the same relative units.

## 3. EXPERIMENTAL RESULTS

*Single-electron capture.* Typical mass spectra of the target particles formed upon the capture of one electron by $^3He^{2+}$ ions are presented in Fig. 1. The spectrum contains clearly displayed peaks corresponding to the formation of singly and doubly charged $C_{60}$ and $C_{70}$ ions during a collision. The {2011} single-electron-capture processes and the {2012} transfer-ionization processes do not cause appreciable fragmentation of the fullerene. Only at large collision energies does the spectrum contain a series of satellites caused by fragmentation of the target molecule with the ejection of light fragments having a mass that is a multiple of two carbon atoms near the $C_{60}^{3+}$ peak, which corresponds to the {2103} process with electron capture and double ionization of the target. However, the fraction of $C_{60}^{3+}$ ions undergoing fragmentation during the capture of one electron with double ionization amounts to less than 10% of the cross section for this process.

The {2001} process is the main process according to the size of the cross section (Fig. 2) over the entire range of collision velocities investigated. The relative role of the $\{201q\}$ transfer-ionization processes, which lead to the formation of multiply charged $C_{60}^{q+}$ ions with $q=2$ and 3, increases in the velocity range investigated from 14 to 45% as the collision velocity increases. This is apparently attributable to the increase in the probability of the capture of one of the inner $\sigma$ electrons of the molecule due to an increase in the velocity of the colliding particles, which, as follows from the structure of a fullerene,[8] leads to the formation of an autoionization state. An increase in the contribution of the cross sections for the transfer-ionization processes as a result of the capture of inner electrons is typical of the interaction of $\alpha$ particles with many-electron targets.[9]

*Double-electron capture.* The mass spectrum of the recoil ions formed as a result of the capture of two electrons by a $He^{2+}$ ion (Fig. 3) is significantly more complicated than the mass spectrum for the capture of one electron. Near each peak corresponds to the formation of $C_{60}^{q+}$ ions with $q>1$ there is a group of satellites due to fragmentation of the molecule. The most intense group corresponds to the capture of two electrons with ionization and with the formation of a series of $C_{60-s}^{3+}$ fragments. The probability of the fragmentation of $C_{60}^{q+}$ ions increases with the magnitude of their charge, and the process involves the detachment of an even number of carbon atoms. It was established from an analysis of mass spectra obtained with different delay times in the channel for detecting the projectile that the processes under consideration are characterized by the formation of a light fragment in an uncharged state. As follows from our work,[10] such a fragmentation scheme is energetically more favorable, since the polarizability coefficient of the heavy fragment is significantly greater than that of the $C_2$ molecule.

In the left-hand part of the mass spectrum there is a group of peaks with values of $M/q$ from 4 to 12, which is associated with the fragmentation of the fullerene. It was shown experimentally that this part of the spectrum corresponds to processes with the fragmentation of the fullerene into two or more charged fragments.

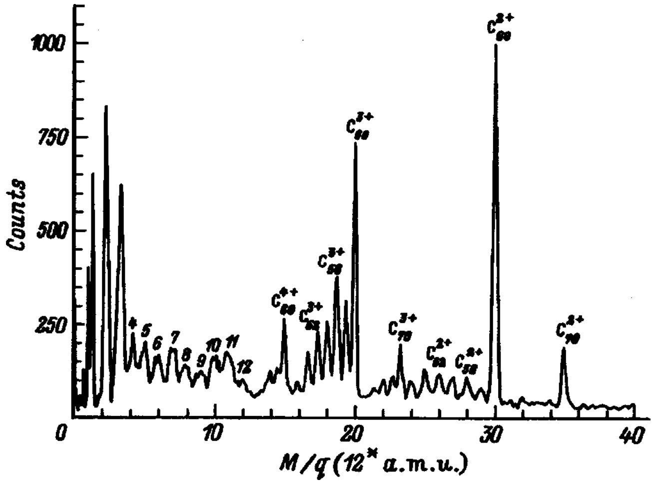

FIG. 3. Mass spectrum of the recoil ions formed in double-electron-capture processes. Energy of the $^3He^{2+}$ ions: $E=6$ keV.

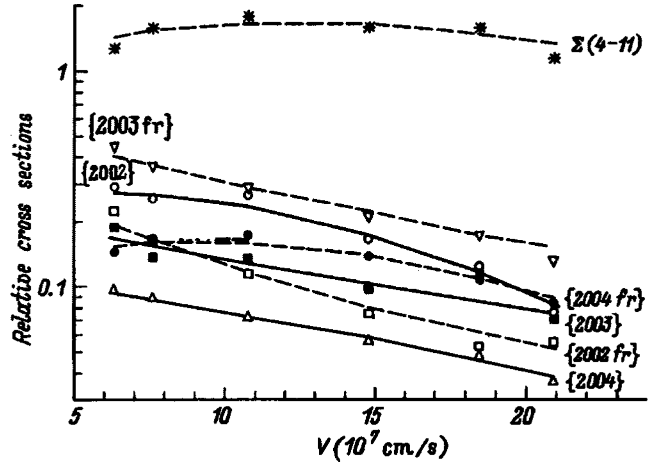

FIG. 4. Relative cross sections for elementary $\{200q\}$ processes due to double-electron capture. Processes without fragmentation: ○—{2002}, ■—{2003}, △—{2004}; processes with fragmentation into a heavy charged $C_{(60-2n)}$ fragment with $n=1-6$ and light neutral fragments: □—{2002 *fr*}; ▽—{2003*fr*}, ●—{2004*fr*}; *—Σ(*4–11*) for processes with fragmentation into two or more charged fragments.

The stability of a multiply charged cluster depends on its size, and it was shown in Ref. 1 that the minimum sizes of the experimentally observable fullerene fragments are as follows: $C_{12}^{2+}$, $C_{20}^{3+}$, $C_{27}^{4+}$, and $C_{36}^{5+}$. Taking into account these data, we can assume that among the peaks observed in the spectrum only the peaks labeled *4* and *5* can be identified as corresponding only to singly charged $C_4^+$ and $C_5^+$ fullerene fragments. On the other hand, it follows from our experimental data that the small peak labeled *12* is caused by the formation of a $C_{60}^{5+}$ ion. Peaks *6–11* can also correspond to multiply charged fullerene fragments.

The cross sections for the elementary processes accompanied by the capture of two electrons by a $He^{2+}$ ion are presented in Fig. 4. Here $\{200qfr\}$ denotes the total cross sections for the processes with the formation of $C_{58}^{q+}-C_{48}^{q+}$ fragments, and Σ(*4–11*) denotes the total cross section for the processes with the formation of several light charged fragments. The contribution of the Σ(*4–11*) channel to double-electron capture is very large over the entire range investigated and increases with the velocity of the colliding particles. Just this finding allows us to assume that the processes with breakup of the fullerene into several fragments, which requires a great expenditure of energy, are caused by the capture of two inner fullerene $\sigma$ electrons by a $He^{2+}$ ion in the ground state of the helium atom. The formation of a recoil ion with two inner vacancies followed by Auger decay of the autoionization state, as well as the direct transfer of the projectile momentum to the carbon atoms, are the most likely reasons for the dominant role of the processes with the formation of several charged fragments.

[1] T. D. Mark and P. Scheier, Nucl. Instrum. Methods Phys. Res. B **98**, 469 (1995).
[2] R. Volpel, G. Hofmann, M. Steidl *et al.*, Phys. Rev. Lett. **71**, 3439 (1993).
[3] B. Walsh, C. L. Cocke, R. Volpel *et al.*, Phys. Rev. Lett. **72**, 1439 (1994).
[4] Z. Wan, J. F. Christian, Y. Basir *et al.*, J. Chem. Phys. **99**, 5858 (1993).
[5] J. F. Christian, Z. Wan, and S. L. Andersen, J. Chem. Phys. **99**, 3468 (1993).
[6] T. LeBrun, H. G. Berry, S. Cheng *et al.*, Nucl. Instrum. Methods Phys. Res. B **98**, 479 (1995).
[7] N. Selberg, A. Barany, C. Biederman *et al.*, Phys. Rev. A **53**, 874 (1996).
[8] U. Thumm, T. Bastug, and B. Frike, Phys. Rev. A **52**, 2955 (1995).
[9] V. V. Afrosimov, A. A. Basalaev, and M. N. Panov, Zh. Tekh. Fiz. **64**(10), 22 (1994) [Tech. Phys. **39**, 989 (1994)].
[10] V. V. Afrosimov, A. A. Basalaev, and M. N. Panov, Zh. Tekh. Fiz. **66**(5), 10 (1996) [Tech. Phys. **41**, 412 (1996)].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Influence of the conduction bandwidth of the substrate on the electronic state of an adatom

S. Yu. Davydov

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

(Submitted February 25, 1997)
Zh. Tekh. Fiz. **68**, 15–19 (April 1998)

The positions of the local and quasilocal levels of an adsorption system and their contributions to the occupation number of an adatom are found within the Anderson Hamiltonian using a simple model of the density of states. The dependence of the occupancy of an adatom on the conduction bandwidth of a metallic substrate is analyzed. © *1998 American Institute of Physics.* [S1063-7842(98)00404-8]

## INTRODUCTION

Among the numerous applications of the Anderson Hamiltonian,[1,2] which was originally proposed for describing magnetic impurities in a metal, its use in the problem of adsorption has become very popular since the papers by Grimley[3] and Newns.[4] In addition, the so-called infinitely-broad-band approximation has been used in many publications, as well as in Anderson's original paper.[1] It is assumed that the quasilevel of an impurity (or, in the terminology of the adsorption problem, the quasilevel of an adatom) is fairly far from the conduction band edge of the metal, i.e., it is so far that the finite value of the conduction bandwidth can be neglected. Such an approach is justified in many problems. Let us, however, imagine the following situation. Let there be a ternary system consisting of a solid metallic substrate, a submonolayer film (i.e., a film in which the relative concentration of adatoms or the coverage $\theta < 1$), and an atom adsorbed on the film. We now assume that the coverage $\theta$ increases. The surface band formed by the overlapping orbitals of the adatoms comprising the film broadens, and its form is modified.[5–9] A similar problem can arise in a two-layer system consisting of a substrate and a submonolayer coating, in which the desorption of an atom from the adsorbed film is studied as a function of the concentration of adatoms in it. In this case the probability of electron exchange between a desorbed atom and coating atoms, or charge exchange, will depend significantly on the conduction bandwidth of the latter. Clearly, it is not possible to describe such a situation in the framework of the broad-band approximation. In addition, as will be shown below, if the conduction bandwidth is finite, when an atom is adsorbed, even if its energy level lies within the band, there is a possibility of the appearance of localized states above and below the band edges, whose presence is completely ignored in the infinitely-broad-band approximation. Thus, the finite value of the conduction bandwidth of the metallic substrate must be taken into account in numerous adsorption problems. The present work is devoted to an investigation of this question.

## MODEL

Let us consider the ground state of an adsorbed one-electron atom, which is described by an Anderson Hamiltonian of the form

$$H=\sum_k \varepsilon_k c_k^+ c_k+\varepsilon_a a^+ a+V\sum_k (c_k^+ a+\text{h.c.}). \tag{1}$$

Here $\varepsilon_k$ is the dispersion law of the metallic electrons of the substrate; $c_k^+(c)$ is the creation (annihilation) operator of electrons in state $k$; $a^+(a)$ is the creation (annihilation) operator of an electron at the $\varepsilon_a$ level; and $V$ is the matrix element describing the hybridization of the states of the atom and the metal. For simplicity, we omitted the spin index and neglected the correlation repulsion of two electrons with opposite spins, assuming it to be so great that the presence of two electrons in one orbital is ruled out.

Let us consider a system consisting of an adatom and a metallic substrate. We assign the density of states of the latter $\rho(\omega)$ by the expression

$$\rho(\omega)=1/D,\ |\omega|<D,\ \rho(\omega)=0,\ |\omega|>D. \tag{2}$$

Here the position of the center of the band is taken as zero energy. It is easy to show (see, for example, Refs. 10 and 11) that the adatom Green's function $G_a$ corresponding to Hamiltonian (1) has the form

$$G_a^{-1}=\omega-\varepsilon_a-\Lambda(\omega)+i\Delta(\omega). \tag{3}$$

Here $\omega$ is the energy, and the broadening $\Delta$ and shift $\Lambda$ of the quasilevel are given, respectively, by the expressions

$$\Delta=\pi V^2/D,\ |\omega|<D,\ \Lambda=0,\ |\omega|>D, \tag{4}$$

$$\Lambda=\frac{V^2}{D}\ln\left|\frac{D+\omega}{D-\omega}\right|. \tag{5}$$

The functions $\Delta(\omega)$ and $\Lambda(\omega)$ are presented in Fig. 1. Generally speaking, the transition to the infinitely-broad-band model ($D\to\infty$) is somewhat artificial: it is assumed that the shift of the atomic level $\Lambda$ can be neglected in this section, while the broadening $\Delta$ must be taken into account.

The density of states in the adatom $\rho_a$ is

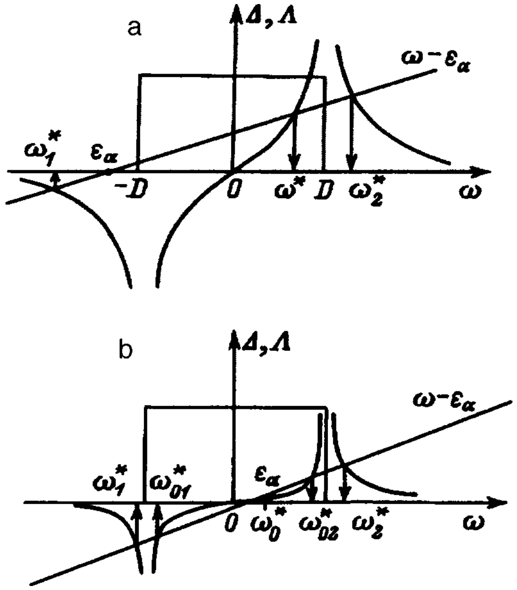

FIG. 1. Broadening function $[\Delta(\omega)]$ and shift function $[\Lambda(\omega)]$ of an atomic level. The conduction band of the metal is located in the range $|\omega|<D$. The roots of Eq. (7) corresponding to local and quasilocal levels are denoted by arrows; a—$2V^2/D^2>1$, b—$2V^2/D^2<1$.

$$\rho_a(\omega)=\frac{1}{\pi}\,\frac{\Delta}{[\omega-\varepsilon_a-\Lambda(\omega)]^2+\Delta^2}. \tag{6}$$

The position of the local or quasilevels of the adsorption system is specified by the equation

$$\omega-\varepsilon_a-\Lambda(\omega)=0. \tag{7}$$

The graphical method for solving Eq. (7) is presented in Fig. 1.

The band contribution $n_b$ to the occupation number of an adatom level is specified by the expression

$$n_b=\frac{V^2}{D}\cdot\int_{-D}^{\varepsilon_f}\frac{d\omega}{[\omega-\varepsilon_a-\Lambda(\omega)]^2+\Delta^2}, \tag{8}$$

where $\varepsilon_f$ is the Fermi energy.

The contribution $n_l$ of a local level lying outside the band of the continuous spectrum of the substrate to the occupation number of the adatom is[11]

$$n_l=\left|1-\frac{\partial\Lambda}{\partial\omega}\right|_{\hat{\omega}}^{-1}, \tag{9}$$

where $\hat{\omega}$ is the root of Eq. (7) corresponding to the local state.

Substituting the expression in (5) therein, we obtain

$$n_e=\left[1+\frac{2V^2}{(\hat{\omega})^2}-D^2\right]. \tag{10}$$

## LOCALIZED AND QUASILOCALIZED LEVELS IN AN ADSORPTION SYSTEM

Let us investigate the solutions of Eq. (7). An analysis shows that for

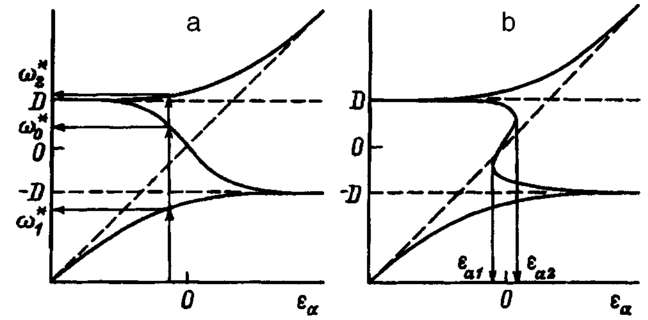

FIG. 2. Dependence of the position of the local and quasilocal levels $\omega_i^*$ on the energy of the atomic level $\varepsilon_a$. $\omega_{01,2}^*=\mp DR$; a—$2V^2/D^2>1$, b—$2V^2/D^2<1$.

$$\frac{2V^2}{D^2}>1 \tag{11}$$

there are two solutions corresponding to the local levels $\omega_{1,2}^*$, one of which (1) lies below the conduction band of the substrate, while the other (2) lies above it, and one solution corresponding to the quasilocal level $\omega^*$ (Fig. 1a). The fulfillment of inequality (11), which corresponds to the so-called case of a ''surface molecule'' (in which the strength of the adatom/substrate-atom interaction exceeds the strength of the interaction between substrate atoms), is simply ruled out in the infinitely-broad-band model. The following circumstance should be noted here. The use of expression (2) for the density of states gives the shift function $\Lambda(\omega)$, which has discontinuities at $\omega=\pm D$ [see Eq. (5) and Fig. 1]. In this case Eq. (1) always has the local solutions $\omega_{1,2}^*$. If, however, a more realistic model of the density of states of the substrate, which does not have a stepwise character, is used, the function $\Lambda(\omega)$ has extrema of finite magnitude at $\omega=\pm D$, and, as a result, the solutions like $\omega_{1,2}^*$ can be absent (see, for example, Refs. 4 and 12). If $\varepsilon_a<0$ and $|\varepsilon_a|\gg D$, then

$$\omega_1^*\cong\frac{\varepsilon_a+\sqrt{\varepsilon_a^2+8V^2}}{2},$$

$$\omega_2^*\cong D\{1+2\exp[-D(D-\varepsilon_a)/V^2]\},$$

$$\omega^*\cong D\{1-2\exp[-D(D-\varepsilon_a)/V^2]\}. \tag{12}$$

When $\varepsilon_a\gg D$, the solution $\omega_2^*$ has the same form as $\omega_1^*$ from (12), but with a plus sign in front of the radical, and $\omega_1^*$ and $\omega^*$ are specified by the second and third formulas in (12), but with a minus sign in front of the right-hand side and with the replacement of $(D-\varepsilon_a)$ by $(D+\varepsilon_a)$ in the numerator of the exponential function.

When $|\varepsilon_a|\ll D$, we have

$$\omega_1^*\cong-D[1+2\exp(-D^2/V^2)],$$

$$\omega_2^*\cong D[1+2\exp(-D^2/V^2)],$$

$$\omega^*\cong-\varepsilon_a/[(2V^2/D^2)-1]. \tag{13}$$

The dependence of the roots of Eq. (7) on the energy of the atomic level $\varepsilon_a$ for (11) in the general case is shown in Fig. 2a. If the inequality

$$\frac{2V^2}{D^2}<1 \tag{14}$$

is satisfied, i.e., if a ‘‘weak-binding’’ regime exists, Eq. (7) has only three roots, which are similar to those described above, or five roots: two ($\omega_1^*$ and $\omega_2^*$) corresponding to levels localized below and above the conduction band and three ($\omega_{01}^*$, $\omega_{02}^*$, and $\omega_0^*$) corresponding to quasilocalized levels located, respectively, in the lower half of the conduction band ($\omega_{01}^*$), in its central region ($\omega_0^*$), and in its upper half ($\omega_{02}^*$). Five solutions exist when the energy of the atomic level $\varepsilon_a$ satisfies the relation

$$\varepsilon_{al}<\varepsilon_a<\varepsilon_{a2}, \tag{15}$$

where

$$\varepsilon_{a1,2}=\mp DR\pm\frac{V^2}{D}\ln\frac{1+R}{1-R},$$

$$R=\sqrt{1-\frac{2V^2}{D^2}}. \tag{16}$$

The dependence of the roots of Eq. (7) on the energy of the atomic level $\varepsilon_a$ for case (14) is presented in Fig. 2b. We note that in the special case in which the energy of the atomic level $\varepsilon_a$ coincides with $\varepsilon_{a1}$ or $\varepsilon_{a2}$ there are only four different solutions, since in these cases either $\omega_{01}^*=\omega_0^*=\overline{\omega}_{01}$ or $\omega_{02}^*=\omega_0^*=\overline{\omega}_{02}$.

## ELECTRON OCCUPANCY OF AN ADATOM

Let us now proceed to an investigation of the occupancy of an adatom. We first consider the contribution $n_l$ of the local states which lie above the band and have the energy $\omega_1^*$. When $\varepsilon_a\to-\infty$, the local level $\omega_1^*\to\varepsilon_a$, and

$$n_l\to(1+2V^2/\varepsilon_a^2)^{-1}, \tag{17}$$

if the inequality (11) holds, and $n_l\to1$ in the case described by (14). When $\varepsilon_a\to0$, we have $\omega_1^*$ defined by (13); therefore,

$$n_l\cong\{1+(2V^2/D^2)[(1+\exp(-D^2/V^2))^2-1]\}^{-1}. \tag{18}$$

When $\varepsilon_a\to\infty$, we have $\omega_1^*\to-D$ and $n_l\to0$.

Let us now consider the contribution of the band states $n_b$ to the occupancy of an adatom.

In an approximation we obtain

$$n_b\cong\frac{1}{\pi}\left[\arctan\frac{D(\omega'+D)}{\pi V^2}-\arctan\frac{D(\omega'-\varepsilon_f)}{\pi V^2}\right]. \tag{19}$$

Here $\omega'=\omega^*$ in the case of the fulfillment of condition (11), and $\omega'=\omega_0^*$, if condition (14) is satisfied, but $\varepsilon_a$ lies outside the interval $(\varepsilon_{a1},\varepsilon_{a2})$ [see (15)]. If the inequality (14) holds and $\varepsilon_a$ lies in the interval $(\varepsilon_{a1},\varepsilon_{a2})$, and if $\varepsilon_f\leqslant\omega_0^*$, $n_b$ can be estimated using the following expression

$$n_b=I_1(\varepsilon_f\leqslant\omega_0^*), \tag{20}$$

where

$$I_1(\varepsilon_f\leqslant\omega_0^*)$$
$$\cong\frac{1}{4\pi\sqrt{F1}}\left[\sin(\alpha/2)\ln\frac{\zeta^2+2\zeta\sqrt{F_1}\cos(\alpha/2)+F_1}{\zeta^2-2\zeta\sqrt{F_1}\cos(\alpha/2)+F_1}\right.$$
$$\left.+2\cos(\alpha/2)\arctan\frac{\zeta^2-F_1}{2\zeta\sqrt{F_1}\sin(\alpha/2)}\right]_{\zeta=-\infty}^{\zeta=E_f-x_{01}}. \tag{21}$$

Here

$$E_a=D\varepsilon_a/\pi V^2,\quad E_f=D\varepsilon_f/\pi V^2,$$

$$d=D^2/\pi V^2,\quad x_{01}=-dR,$$

$$F_1^2=f_{01}^2+1,\ f_{01}=-dR-E_a+\frac{1}{\pi}\ln\frac{1+R}{1-R},$$

$$\sin(\alpha/2)=[(1+f_{01}/F_1)/2]^{1/2}. \tag{22}$$

In (21) the lower integration limit is taken equal to $-\infty$, rather than to the exact value $\zeta=-d-x_{01}$. This was done to take into account the fact that the integrand in (8), i.e., the density of states in the adatom, vanishes as $\omega\to-D$, because the shift function $\Lambda$ goes to infinity.

When $\varepsilon_f>\omega_0^*$,

$$n_b=[I_1(\varepsilon_f=\omega_0^*)+I_2(\varepsilon_f>\omega_0^*)]. \tag{23}$$

Here $I_2(\varepsilon_f>\omega_0^*)$ is given by (21) after the replacement of $F_1$ by $F_2$, $f_{01}$ by $f_{02}$, and $x_{01}$ by $x_{02}$, where

$$F_2^2=f_{02}^2+1,\ f_{02}=dR-E_a-\frac{1}{\pi}\ln\frac{1+R}{1-R},\ x_{02}=-x_{01},$$

$$\sin(\alpha/2)=[(1-f_{02}/F_2)]^{1/2}, \tag{24}$$

the upper integration limit is $\zeta=E_f-x_{02}$, and the lower integration limit is $\zeta=\omega_0^*-x_{02}$. If $E_f\to D$, the upper limit should be replaced by infinity. We note that (19), (21), and (23) give overestimated values of $n_b$.

## DEPENDENCE OF THE OCCUPANCY OF AN ADATOM ON THE CONDUCTION BANDWIDTH

The results of the calculations of the contribution of the local state $n_l$ and the contribution of the band $n_b$ are presented in Figs. 3–5, respectively. It follows from Fig. 3 that $n_l$ decreases with increasing $D$. The slope of the $n_l(D)$ curve is proportional to the proximity of the energy of the local state $\omega_1^*$ to the lower conduction band edge $(-D)$ [when $D=1$, in Fig. 3a we have $\omega_1^*=-3.16$ for curve *1*, $\omega_1^*=-1.81$ for curve *2*, $\omega_1^*=-1.03$ for curve *3*, and $\omega_1^*\cong-D$, while $n_l$ vanishes for curves *4* and *5*; see Eq. (13)]. When $\omega_1^*$ is at a distance from the lower conduction band edge exceeding $D$, the value of $\omega_1^*$ is specified by (12) and, therefore, depends very weakly on $D$ (curve *1* in Figs. 3a and 3b).

Dependences of the band contribution $n_b$ to the occupancy of an adatom as a function of $D$ are presented in Figs. 4 and 5 for various values of the occupancy of the band $(\varepsilon_f)$. We first consider the weak-binding regime (Fig. 4). When $\varepsilon_a\gg D$, the energy of the quasilevel $\omega^*\cong-D$ (Fig. 2), and from (19) we obtain

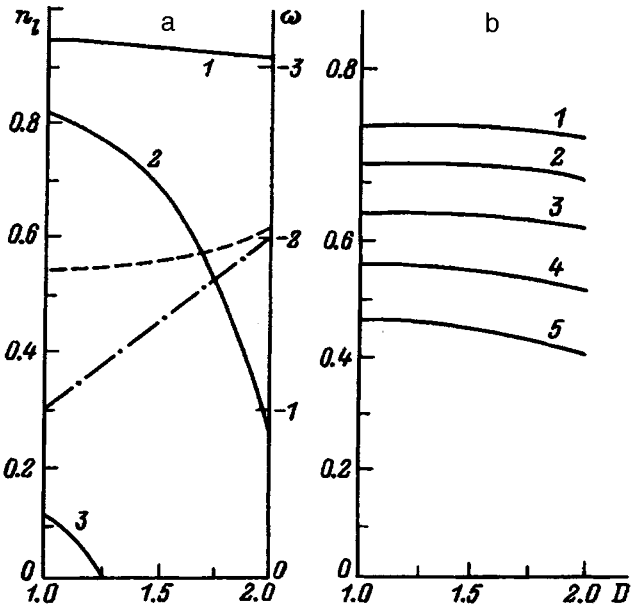

FIG. 3. Dependence of the contribution $n_l$ of a local level lying below the conduction band to the occupation number on the conduction bandwidth $D$ (solid lines), as well as variation of the position of the local level (dashed line) and the position of the lower conduction band edge (dot-dashed line). $\varepsilon_a=-3$ (*1*), $-1.5$ (*2*), 0 (*3*), 1.5 (*4*), 3 (*5*); $V=0.5$ (a) and 2 (b); in case (a) the values of $n_l$ for $\varepsilon_a=1.5$ and 3 are very small and are not shown in the figure.

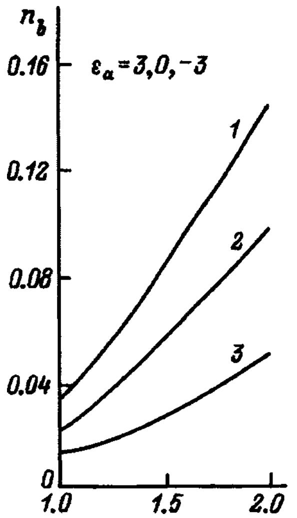

FIG. 5. Same as in Fig. 4, but for $V=2$. The results of the calculation for $\varepsilon_a=3$, 0, and $-3$ coincide on the scales of the figure.

$$n_b \cong \frac{1}{\pi}\arctan\frac{D(D+\varepsilon_f)}{\pi V^2}. \qquad (25)$$

As $D$ and $\varepsilon_f$ increase, the value of $n_b$ also increases, as is observed in Fig. 4a. In the case of $\varepsilon_a<0$ and $|\varepsilon_a|\gg D$, where $\omega^*\cong D$, the descending $n_b(D)$ curve presented in Fig. 4c is determined by the expression

$$n_0 \cong \frac{1}{\pi}\left[\arctan\frac{2D^2}{\pi V^2}-\arctan\frac{D(D-\varepsilon_f)}{\pi V^2}\right]. \qquad (26)$$

As $D\to\infty$, $n_b\to 0$, and $\partial n_b/\partial\varepsilon_f \to V^2/[D(D-\varepsilon_f)^2]$, i.e., the slope of the $n_b(D)$ curve increases as $\varepsilon_f$ increases (Fig. 4c). A change in the character of $n_b(D)$ from an increasing function to a decreasing function as $\varepsilon_a$ varies from $-\infty$ to $+\infty$ occurs when the energy of the atomic term crosses the conduction band of the metallic substrate.

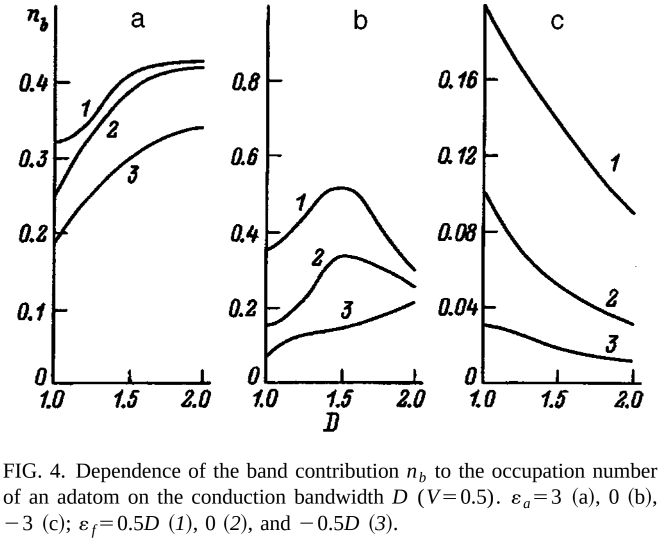

FIG. 4. Dependence of the band contribution $n_b$ to the occupation number of an adatom on the conduction bandwidth $D$ $(V=0.5)$. $\varepsilon_a=3$ (a), 0 (b), $-3$ (c); $\varepsilon_f=0.5D$ (*1*), 0 (*2*), and $-0.5D$ (*3*).

The $n_b(D)$ curve for the surface-molecule regime is presented in Fig. 5. The values of $n_b$ for the three values of the energy of the atomic level $\varepsilon_a=3$, 0, and $-3$ are indistinguishable on the scales of the figure. In fact, when $\pi V^2 \gg D^2$, instead of (19) we obtain

$$n_b \simeq \frac{D(D+\varepsilon_f)}{\pi^2 V^2}, \qquad (27)$$

i.e., the value of the occupancy does not depend on the position of the quasilevel.

Thus, the calculations show that a change in the conduction bandwidth of the substrate $\delta D$ leads to a significant change in the occupancy of an adatom $\delta n_a$. Moreover, the effect of broadening (or narrowing) of the band can exceed the effect of a change in its occupancy $(\delta\varepsilon_f)$. If we are referring to a three-layer system consisting of a solid substrate with a submonolayer $(\theta\leqslant 1)$ metal film deposited on it and an atom adsorbed on the surface of the film, the following can be noted. First, within both the tight-binding model[13,14] and the free-electron approximation[15] the width of the surface band of the film is proportional to $a^{-2}$, where $a=a(\theta)$ is the distance between nearest neighbors, which depends on the relative concentration of atoms in the film $\theta=N_m/N_m(\mathrm{ML})$, $N_m(\mathrm{ML})$ being the surface concentration of atoms in a monolayer. Since $N_m=a^{-2}$ and $N_m(\mathrm{ML})=d^{-2}$, where $d$ is the distance between nearest neighbors in a monolayer (ML), we have

$$D(\theta)\propto D(\mathrm{ML})\cdot\theta. \qquad (28)$$

When there is a dipole-dipole interaction between the adatoms, the concentration shift of the Fermi level of the adsorption system $\delta\varepsilon_f$ equals $\Phi\cdot\theta^{3/2}(n_m-1)$, where $\Phi=2e^2l^2A/d^3$ [$l$ is half of length of the dipole formed by an adsorbed ion and its image in the metallic substrate, $d$ is the distance between nearest-neighbor adatoms in a monolayer film, $n_m$ is the occupancy of an adatom, and $A \cong 9$ (Refs. 5 and 16)]. In the case of alkali metal atoms, if $l$ is assumed to be equal to half of the sum of their ionic and atomic radii, and the distance between nearest neighbors in a crystal of the alkali metal is taken as $d$, we obtain $\Phi \cong 10$ eV. Thus, $\Phi$ and $D$ are of the same order of magnitude. Therefore, the linear variation of the conduction bandwidth can play the decisive role, i.e., it can influence charge exchange more strongly than can variation of the occupancy.

We thank N. D. Potekhina for some valuable comments. This work was performed as part of the Surface Atomic Structures Program.

[1] P. W. Anderson, Phys. Rev. **124**, 41 (1961).

[2] P. W. Anderson, Science **201**, 307 (1978) [Usp. Fiz. Nauk **127**, 19 (1979)].

[3] T. B. Grimley, Proc. Phys. Soc. London **90**, 751 (1967).

[4] D. M. Newns, Phys. Rev. **178**, 1123 (1969).

[5] S. Yu. Davydov, Fiz. Tverd. Tela (Leningrad) **19**, 3376 (1977) [Sov. Phys. Solid State **19**, 1971 (1977)].

[6] S. Yu. Davydov, Fiz. Tverd. Tela (Leningrad) **20**, 1752 (1978) [Sov. Phys. Solid State **20**, 1013 (1978)].

[7] S. Yu. Davydov, Fiz. Tverd. Tela (Leningrad) **20**, 1998 (1978) [Sov. Phys. Solid State **20**, 1153 (1978)].

[8] S. Yu. Davydov, Fiz. Met. Metalloved. **47**, 481 (1979).

[9] S. Yu. Davydov, Poverkhnost', No. 8, 17 (1991).

[10] C. Kittel, *Quantum Theory of Solids*, Wiley, New York (1963), Chap. 18 [Russian trans., Nauka, Moscow (1967), 491 pp.].

[11] F. D. M. Haldane and P. W. Anderson, Phys. Rev. B **13**, 2553 (1976).

[12] *Theory of Chemisorption*, J. R. Smith (ed.), Springer-Verlag, Berlin (1980) [Russian trans., Mir, Moscow (1983), 333 pp.].

[13] W. A. Harrison, *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond*, W. H. Freeman, San Francisco (1979) [Russian trans., Mir, Moscow (1983), 382 pp.].

[14] W. A. Harrison, Phys. Rev. B **27**, 3592 (1983).

[15] C. Kittel, *Introduction to Solid State Physics*, 5th ed., Wiley, New York (1976) [Russian trans., Mir, Moscow (1978), 792 pp.].

[16] J. P. Muscat and D. M. Newns, J. Phys. C: Solid State Phys. **7**, 2630 (1974).

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Capillary oscillations of a charged viscous spheroidal droplet

S. O. Shiryaeva

*Yaroslavl State University, 150000 Yaroslavl, Russia*

(Submitted August 23, 1996)
Zh. Tekh. Fiz. **68**, 20–27 (April 1998)

Dispersion relations are derived for the capillary oscillations of a charged viscous spheroidal droplet by scalarization within perturbation theory using an expansion in two small parameters, viz., the magnitude of the perturbation of the spheroidal surface as a result of thermal fluctuations and the magnitude of the deviation of the equilibrium spheroidal droplet shape from a spherical shape. It is shown analytically that the motion spectrum of the liquid consists of two components that interact in the linear theory: toroidal vortex motion and poloidal potential motions. A numerical analysis reveals that the instability growth rates of the higher modes of a highly charged droplet increase with enhancement of the degree of spheroidal strain and decrease rapidly as the viscosity of the liquid increases. © *1998 American Institute of Physics.* [S1063-7842(98)00504-2]

The investigation of the capillary oscillations of a charged droplet of a viscous liquid and its stability against its self-charge is of interest for geophysics, technical physics, chemical engineering, etc. The experimental and technical investigation of this object has been the subject of numerous publications.[1] Nevertheless, many questions related to this problem have been studied to only a small extent. This refers primarily to the physical mechanism for the development of the instability of a charged droplet with respect to its self-charge. Rayleigh showed[2] that the first of the infinite set of capillary modes of a highly charged droplet to experience instability is the fundamental mode, which is proportional to the Legendre polynomial $P_2(\cos\beta)$. The development of instability causes the droplet to stretch out into a figure that is spheroidal in the linear approximation with respect to the square of the eccentricity.[3] This naturally raises the question of how alteration of the shape of an originally spherical droplet influences the laws governing the realization of the instability of the higher capillary modes. The corresponding problem was formulated in Ref. 3, but it was solved only on a qualitative level of rigor. Nevertheless, it was ascertained that the critical conditions for the realization of the instability of capillary modes higher than the fundamental mode ease, as the eccentricity of a highly charged spheroidal droplet increases. This phenomenon underlies the qualitative physical mechanism for the realization of the instability of a charged droplet proposed in Ref. 3. In Ref. 4 an attempt was made to find the spectrum of capillary oscillations and damping decrements for a charged spheroidal droplet of a viscous liquid in a one-dimensional electrostatic field by writing out the Lagrangian of the oscillatory system with consideration of the dissipation. The spectra of the capillary oscillations and their damping decrements for a spheroidal viscous droplet were calculated analytically in Refs. 5 and 6.

The investigation of the temporal evolution of the amplitude of the fundamental mode of a droplet which is unstable toward its self-charge in Ref. 7 called for a search for a rigorous solution of the problem of the influence of the spheroidal shape of a charged droplet on the growth rates of the instability of its various modes toward its self-charge.

**1.** We shall solve the problem of the axisymmetric capillary oscillations of a charged prolate spheroidal droplet of a viscous, ideally conducting liquid, assuming that the spheroidal shape of the droplet is caused either by virtual deformation or by the action of some stationary (not dependent on the time and the droplet shape) extraneous forces of nonelectrical nature (see the Appendix). We present the solution in dimensionless variables, setting the radius $R$ of the original spherical droplet, the density $\rho$, and the surface tension $\sigma$ equal to unity: $R=1$, $\rho=1$, $\sigma=1$. Then, as units of measure of the distance, time, charge, pressure, velocity, and viscosity, we obtain the characteristic values $r_*=R$, $t_*=R^{3/2}\rho^{1/2}\sigma^{-1/2}$, $Q_*=R^{3/2}\sigma^{1/2}$, $p_*=R^{-1}\sigma$, $u_*=R^{-1/2}\rho^{-1/2}\sigma^{1/2}$, and $\nu_*=R^{1/2}\rho^{-1/2}\sigma^{1/2}$.

The equation of the surface of a prolate spheroid perturbed by capillary wave motion in spherical coordinates in the linear approximation with respect to $e^2$, i.e., the square of the eccentricity of the spheroid, has the form

$$r=r(\theta)+\xi(\theta,t)\approx 1+e^2h(\theta)+\xi(\theta,t);$$

$$r(\theta)=\frac{(1-e^2)^{1/6}}{(1-e^2\cos^2\theta)^{1/2}};\quad h(\theta)=\frac{1}{6}(3\cos^2\theta-1),\tag{1.1}$$

which describes an equilibrium spheroidal droplet surface perturbed by capillary oscillations that occur because of the thermal motion of the molecules and have an amplitude $\sim\sqrt{kT/\sigma}$ (here $k$ is Boltzmann's constant, and $T$ is the absolute temperature). We also note that for most liquids the amplitude of such thermal capillary oscillations is of the order of an angstrom unit.

The ensuing analysis is performed within perturbation theory using an expansion in the small parameters $\xi$ and $e^2$ (in the order indicated) to terms $\sim\xi$ and $\sim e^2\xi$, i.e., in the linear approximation with respect to $\xi$. We note that the small parameters $e^2$ and $\xi$ are independent, but it is assumed that $e^2 \gg \xi$. For this reason, it would appear that, if we retain the terms $\sim e^2\xi$, we should also take into account the terms $\sim e^4$. However, as will be seen below, only the terms $\sim \xi$ and $\sim e^2\xi$ make contributions to the dispersion relation sought, while the terms $\sim e^2$ and $\sim e^4$ vanish when the kinematic boundary condition (which contains a partial derivative with respect to time) is taken into account. Therefore, retention of the terms $\sim e^4$ in the calculations would only lead to an unjustified increase in the unwieldiness of the mathematical manipulations.

The time dependences of the velocity field $\mathbf{U}(\mathbf{r},t)$, the pressure field $p(\mathbf{U},t)$, and the perturbation $\xi(\theta,t)$ are assumed to be exponential, i.e., $\sim \exp(st)$.

We write out the system of hydrodynamic equations describing the motion of a viscous liquid in a droplet that is caused by a small perturbation $\xi(\theta,t)$ of the shape of its equilibrium surface and is therefore characterized by a velocity field $\mathbf{U}(\mathbf{r},t)$, which is proportional to the small parameter $\xi$. The system consists of a linearized Navier-Stokes equation and the incompressibility condition

$$\frac{\partial \mathbf{U}}{\partial t} = -\boldsymbol{\nabla}(p(\mathbf{U},t)) + \nu\nabla^2\mathbf{U}; \tag{1.2}$$

$$\operatorname{div} \mathbf{U} = 0, \tag{1.3}$$

where $p(\mathbf{U},t)$ is an addition to the pressure within the liquid, which is first-order with respect to $U$ (with respect to $\xi$).

The following boundary conditions should hold on the droplet surface (1.1):

$$\frac{\partial F(\mathbf{r},t)}{\partial t} + \mathbf{U}\cdot\boldsymbol{\nabla}F(\mathbf{r},t) = 0;$$

$$F(\mathbf{r},t) \equiv r - [1 + e^2 h(\theta) + \xi(\theta,t)] = 0; \tag{1.4}$$

$$\boldsymbol{\tau}\cdot(\mathbf{n}\cdot\boldsymbol{\nabla})\mathbf{U} + \mathbf{n}\cdot(\boldsymbol{\tau}\cdot\boldsymbol{\nabla})\mathbf{U} = 0; \tag{1.5}$$

$$-p(\mathbf{U},t) + 2\nu\mathbf{n}\cdot(\mathbf{n}\cdot\boldsymbol{\nabla})\mathbf{U} - p_E(\xi) + p_\sigma(\xi) = 0, \tag{1.6}$$

where $\mathbf{n}$ and $\boldsymbol{\tau}$ are the vectors of a normal and a tangent to the surface, and $p_E(\xi)$ and $p_\sigma(\xi)$ are additions to the pressure of the electric field and the pressure of the surface tension forces that result from the perturbation of the equilibrium surface of the spheroid and are first-order with respect to $\xi$.

We also require fulfillment of the condition for constancy of the droplet volume

$$\int_0^\pi \xi(\theta,t)\sin\theta\, d\theta = 0 \tag{1.7}$$

and the condition for a stationary center of mass

$$\int_0^{2\pi}\int_0^\pi \xi(\theta,t)\mathbf{e}_r \sin\theta\, d\theta\, d\varphi = 0, \tag{1.8}$$

where $\mathbf{e}_r$ is the radial unit vector of the spherical coordinate system.

We note that in solving the problem in the first-order approximation with respect to the small parameters $U$ and $\xi$, it is sufficient to take the boundary conditions (1.4)–(1.6) on the unperturbed surface of the spheroidal droplet $r = r(\theta) \approx 1 + e^2 h(\theta)$, since they consist of terms that are linear in $U$ and $\xi$.

The system of vector equations (1.2) and (1.3) with the boundary conditions (1.4)–(1.6) and the additional conditions (1.7) and (1.8) comprises the complete mathematical formulation of the problem posed.

**2.** We solve the system of equations (1.2) and (1.3) by scalarization, which was described in detail in Ref. 8. For this purpose, we represent the velocity field $\mathbf{U}(\mathbf{r},t)$ in the form of the expansion

$$\mathbf{U}(\mathbf{r},t) = \hat{\mathbf{N}}_1\Phi_1(\mathbf{r},t) + \hat{\mathbf{N}}_2\Phi_2(\mathbf{r},t) + \hat{\mathbf{N}}_3\Phi_3(\mathbf{r},t), \tag{2.1}$$

where the $\Phi_i(\mathbf{r},t)$ are scalar functions, and the $\hat{\mathbf{N}}_i$ are differential vector operators, which have the following forms in a spherical coordinate system:

$$\hat{\mathbf{N}}_1 \equiv \boldsymbol{\nabla}; \quad \hat{\mathbf{N}}_2 \equiv \boldsymbol{\nabla}\times\mathbf{r}; \quad \hat{\mathbf{N}}_3 \equiv \boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{r}). \tag{2.2}$$

Substituting the expansion (2.1) into (1.2) and (1.3) and using the properties of operators (2.2) $\hat{\mathbf{N}}_k\nabla^2 = \nabla^2\hat{\mathbf{N}}_k$, $\hat{\mathbf{N}}_j^+ \cdot \hat{\mathbf{N}}_k = 0$ (for $k \neq j$) and $\hat{\mathbf{N}}_1^+ = -\hat{\mathbf{N}}_1$, we can easily transform the system of vector equations (1.2) and (1.3) for $\mathbf{U}(\mathbf{r},t)$ and $p(\mathbf{U},t)$ into a system of scalar equations for the scalar functions $\Phi_k(\mathbf{r},t)$:

$$\nabla^2\Phi_k(\mathbf{r},t) - \frac{s}{\nu}(1-\delta_{k1})\Phi_k(\mathbf{r},t) = 0 \quad (k = 1, 2, 3);$$

$$p(\mathbf{U},t) = -s\Phi_1(\mathbf{r},t). \tag{2.3}$$

Here the superscript plus sign denotes a Hermitian conjugate, and $\delta_{kj}$ is the Kronecker delta.

The boundary conditions (1.4)–(1.6) should also be rewritten in terms of the scalar functions $\Phi_k(\mathbf{r},t)$. Using (2.1) and (2.2) and writing out the representations of the components of the velocity vector $\mathbf{U}(\mathbf{r},t)$ in terms of the functions $\Phi_k(\mathbf{r},t)$, after some simple, but cumbersome mathematical transformations[6] we can obtain the kinematic boundary condition for the axisymmetric case in the linear approximation with respect to $e^2$

$$\begin{aligned} r = 1: \quad & s\xi - \left\{\left[\frac{\partial\Phi_1}{\partial r} - \nabla_\Omega^2\left(\frac{\Phi_3}{r}\right)\right] - e^2\left[h(\theta)\frac{\partial}{\partial r}\right.\right. \\ & \times\left(\frac{\partial\Phi_1}{\partial r} - \nabla_\Omega^2\left(\frac{\Phi_3}{r}\right)\right) + \sin\theta\cos\theta\frac{\partial}{\partial\theta} \\ & \left.\left.\times\left(\frac{\Phi_1}{r} + \frac{1}{r}\frac{\partial}{\partial r}(r\Phi_3)\right)\right]\right\} = 0; \end{aligned} \tag{2.4}$$

the dynamic boundary condition for the tangential components of the stress tensor (1.5) breaks up into two conditions in accordance with the need to select two mutually perpendicular vectors as unit vectors that are tangential to the surface

$$\begin{aligned} r=1:\quad &\biggl\{\nabla^2_\Omega\biggl[\frac{\partial\Phi_2}{\partial r}-\frac{\Phi_2}{r}\biggr]+e^2\biggl[(\cos^2\theta-1)\\ &\times\nabla^2_\Omega\biggl(\frac{\Phi_2}{r}\biggr)+h(\theta)\nabla^2_\Omega\frac{\partial}{\partial r}\biggl(\frac{\partial\Phi_2}{\partial r}-\frac{\Phi_2}{r}\biggr)\\ &+\sin\theta\cos\theta\frac{\partial}{\partial\theta}(4+\nabla^2_\Omega)\biggl(\frac{\Phi_2}{r}\biggr)\biggr]\biggr\}=0; \end{aligned} \tag{2.5}$$

$$\begin{aligned} &\biggl\{\nabla^2_\Omega\biggl[2\frac{\partial}{\partial r}\biggl(\frac{\Phi_1}{r}\biggr)+\frac{\partial^2\Phi_3}{\partial r^2}-(2+\nabla^2_\Omega)\biggl(\frac{\Phi_3}{r^2}\biggr)\biggr]\\ &-2e^2\biggl[(3\cos^2\theta-1)\biggl[\frac{\partial^2\Phi_1}{\partial r^2}-\frac{1}{r}\frac{\partial\Phi_1}{\partial r}\biggr]\\ &-(2\cos^2\theta-1)\nabla^2_\Omega\biggl(\frac{\Phi_1}{r^2}\biggr)-(5\cos^2\theta-2)\\ &\times\nabla^2_\Omega\biggl(\frac{1}{r}\frac{\partial\Phi_3}{\partial r}\biggr)+(4\cos^2\theta-1)\nabla^2_\Omega\biggl(\frac{\Phi_3}{r^2}\biggr)-\frac{1}{2}h(\theta)\\ &\times\nabla^2_\Omega\frac{\partial}{\partial r}\biggl[2\frac{\partial}{\partial r}\biggl(\frac{\Phi_1}{r}\biggr)+\frac{\partial^2\Phi_3}{\partial r^2}-(2+\nabla^2_\Omega)\biggl(\frac{\Phi_3}{r_2}\biggr)\biggr]\\ &+\sin\theta\cos\theta\frac{\partial}{\partial\theta}\biggl[\frac{\partial^2\Phi_1}{\partial r^2}-\frac{1}{r}\frac{\partial\Phi_1}{\partial r}-(2+\nabla^2_\Omega)\biggl(\frac{\Phi_1}{r^2}\biggr)\\ &-2(1+\nabla^2_\Omega)\biggl(\frac{1}{r}\frac{\partial\Phi_3}{\partial r}\biggr)-(2-\nabla^2_\Omega)\biggl(\frac{\Phi_3}{r^2}\biggr)\biggr]\biggr]\biggr\}=0; \end{aligned} \tag{2.6}$$

and the dynamic boundary condition for the normal component of the stress tensor is

$$\begin{aligned} r=1:\quad &\biggl\{s\biggl[\Phi_1+e^2h(\theta)\frac{\partial\Phi_1}{\partial r}\biggr]\\ &+2\nu\biggl[\biggl[\frac{\partial^2\Phi_1}{\partial r^2}-\nabla^2_\Omega\biggl(\frac{\partial}{\partial r}\biggl(\frac{\Phi_3}{r}\biggr)\biggr)\biggr]+e^2h(\theta)\frac{\partial}{\partial r}\\ &\times\biggl[\frac{\partial^2\Phi_1}{\partial r^2}-\nabla^2_\Omega\biggl(\frac{\partial}{\partial r}\biggl(\frac{\Phi_3}{r}\biggr)\biggr)\biggr]+\sin\theta\cos\theta\frac{\partial}{\partial\theta}\\ &\times\biggl[\frac{2}{r}\frac{\partial\Phi_1}{\partial r}-\frac{\Phi_1}{r^2}+\frac{\partial^2\Phi_3}{\partial r^2}+\frac{\partial}{\partial r}\biggl(\frac{\Phi_3}{r}\biggr)-\nabla^2_\Omega\biggl(\frac{\Phi_3}{r^2}\biggr)\biggr]\biggr]\\ &-p_E(\xi)+p_\sigma(\xi)\biggr\}=0; \end{aligned}$$

$$\nabla^2_\Omega\equiv\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\biggl(\sin\theta\frac{\partial}{\partial\theta}\biggr). \tag{2.7}$$

System of equations (2.3) with conditions (2.4)–(2.7) comprises the mathematical formulation of the problem being solved in a scalarized form. The solutions of Eqs. (2.3) in the spherical coordinate system which are regular at the point $r=0$ have the form

$$\Phi_1(\mathbf{r},t)=\sum_{m=2}^{\infty}C^1_m r^m Y^0_m(\theta)\exp(st);$$

$$\Phi_{2,3}(\mathbf{r},t)=\sum_{m=2}^{\infty}C^{2,3}_m i_m\biggl(\sqrt{\frac{s}{\nu}}\,r\biggr)Y^0_m(\theta)\exp(st), \tag{2.8}$$

where $i_m(x)$ is a modified spherical Bessel function.

It is also convenient to represent $\xi(\theta,t)$ in the form of an expansion in spherical functions

$$\xi(\theta,t)=\sum_{m=2}^{\infty}Z_m Y^0_m(\theta)\exp(st), \tag{2.9}$$

where the $Z_m$ are constants.

In expansions (2.8) and (2.9) the minimum value of the index $m=2$ because of the additional conditions (1.7) and (1.8).

**3.** The relation between the unknown constants $C^k_m$ ($k=1, 2, 3$) in expansions (2.8) and the coefficients $Z_m$ in the expansion of the perturbation function $\xi(\theta,t)$, as well as the expression for the complex frequency of the capillary oscillations $s$ are determined from the boundary conditions (2.4)–(2.7).

It is easy to see that in the case of axisymmetric oscillations the boundary condition (2.5) for $\Phi_2(\mathbf{r},t)$ is completely autonomous, i.e., it does not contain any other unknown function. Therefore, we describe its treatment, concentrating our attention on the boundary conditions (2.4), (2.6), and (2.7) for $\Phi_1(\mathbf{r},t)$, $\Phi_3(\mathbf{r},t)$, and $\xi(\theta,t)$.

Substituting solutions (2.8) and (2.9) into conditions (2.4), (2.6), and (2.7), using the recurrence relations for cylindrical spherical functions, representing the products $\cos^2\theta\cdot Y^0_m(\theta)$, $h(\theta)\cdot Y^0_m(\theta)$, and

$$\sin\theta\cdot\cos\theta\frac{\partial Y^0_m(\theta)}{\partial\theta}$$

in the form of expansions in spherical functions,[9] and neglecting the interaction between different modes, in the linear approximation with respect to $e^2$ from the kinematic boundary condition we can obtain (2.4)

$$\begin{aligned} &C^1_m\{m+e^2[m(m-1)-3]\varkappa_m\}+D^3_m\\ &\times\biggl\{m(m+1)+e^2\varkappa_m\biggl[\sqrt{\frac{s}{\nu}}\,f_m\biggl(\sqrt{\frac{s}{\nu}}\biggr)(m(m+1)-3)\\ &+(m+1)(m(m-1)-3)\biggr]\biggr\}-sZ_m=0, \end{aligned} \tag{3.1}$$

from the dynamic boundary condition for the tangential component of the stress tensor (2.6) we can obtain

$$\begin{aligned}&C_m^1\Bigl\{2(m-1)+e^2\varkappa_m 2(m-1)\Bigl[(m-2)-\frac{9}{m(m+1)}\Bigr]\Bigr\}\\&+D_m^3\Bigl\{\Bigl[\frac{s}{\nu}+2(m^2-1)-2\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)\Bigr]\\&+e^2\varkappa_m\Bigl[\frac{s}{\nu}(m-2)+2(m^2-1)\\&\times\Bigl((m-2)-\frac{9}{m(m+1)}\Bigr)+\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)\\&\times\Bigl(\frac{s}{\nu}+2(m^2+m+2)+\frac{18}{m(m+1)}\Bigr)\Bigr]\Bigr\}=0,\end{aligned}\tag{3.2}$$

and from the dynamic boundary condition for the normal component of the stress tensor (2.7) we can obtain

$$\begin{aligned}&C_m^1\Bigl\{\Bigl[\frac{s}{\nu}+2m(m-1)\Bigr]+e^2\varkappa_m\Bigl[\frac{s}{\nu}m\\&+2(m^3-3m^2-4m+3)\Bigr]\Bigr\}+D_m^3\Bigl\{2m(m+1)\\&\times\Bigl[(m-1)+\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)\Bigr]+e^2\varkappa_m 2\\&\times\Bigl[\frac{s}{\nu}(m(m+1)-3)+(m+1)(m^3-3m^2-4m+3)\\&-\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)(4m(m+1)-3)\Bigr]\Bigr\}\\&+Z_m\Bigl\{(m-1)(m+2)\,\frac{\alpha_m}{\nu}\\&+e^2\varkappa_m\Bigl[(m+2)(m-4)\frac{\alpha_m}{\nu}-3\frac{m^2}{\nu}\Bigr]\Bigr\}=0.\end{aligned}\tag{3.3}$$

In expressions (3.1)–(3.3) we introduced the following notation:

$$\varkappa_m\equiv\frac{m(m+1)}{3(2m-1)(2m+3)};\quad D_m^3\equiv C_m^3 i_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr);$$

$$f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)\equiv\frac{i_{m+1}\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)}{i_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)};\quad \alpha_m\equiv 1-\frac{W}{(m+2)};\quad W\equiv\frac{Q^2}{4\pi}.$$

Equation (3.3) was written using expressions for the first-order additions to the pressure of the electric forces $p_E(\xi)$ and the pressure of the surface-tension forces $p_\sigma(\xi)$ with respect to the small parameter $\xi$, which have the following forms in the axisymmetric case with neglect of the interaction between different modes[7,10]

$$\begin{aligned}p_E(\xi)&\approx\frac{Q^2}{4\pi}\sum_{m=2}^{\infty}\{[(m-1)+e^2\\&\quad\times(m-4)\varkappa_m]Z_m\}Y_m^0(\theta)\exp(st);\\p_\sigma(\xi)&\approx\sum_{m=2}^{\infty}\{[(m-1)(m+2)-e^2 2\\&\quad\times(m^2+m+4)\varkappa_m]Z_m\}Y_m^0(\theta)\exp(st).\end{aligned}$$

The condition that the determinant of the system of equations (3.1)–(3.3) vanish yields the dispersion relation of the problem being solved in the linear approximation with respect to $e^2$

$$\begin{aligned}&s[s^2+m(m-1)(m+2)\alpha_m]+2\nu\Bigl[s^2(m-1)(2m+1)\\&-\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)(s^2+m(m-1)(m+2)\alpha_m)\Bigr]\\&-4\nu^2 m(m-1)(m+2)s\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)\\&+e^2\varkappa_m\Bigl\{s\Bigl[(s^2+m(m-1)(m+2)\alpha_m)\\&\times\Bigl(2(m-1)+\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)\Bigr)\\&-3((2m-1)(m+2)\alpha_m+m^3)\Bigr]\\&+2\nu\Bigl[s^2\Bigl(2m^3-8m^2+4m-10+\frac{9}{m}\Bigr)+\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)\\&\times\Bigl(s^2\Bigl(2m^2-m+2+\frac{9}{m(m+1)}\Bigr)+3\Bigl(2m^2-m+3\\&-\frac{6}{(m+1)}\Bigr)(m+2)\alpha_m+3m^3\Bigr)\Bigr]+4\nu^2 s\sqrt{\frac{s}{\nu}}\,f_m\Bigl(\sqrt{\frac{s}{\nu}}\Bigr)\\&\times\Bigl[5m^3+5m^2+2m-\frac{18}{(m+1)}\Bigr]\Bigr\}=0.\end{aligned}\tag{3.4}$$

The general form of the motions of the liquid realized (the branches of the dispersion relation) in the form of plots of Re $s=$Re $s(W)$ and Im $s=$Im $s(W)$ obtained by numerical calculations using (3.4) for $m=2$, $\nu=0.03$, and $e^2=0.2$ is presented in Fig. 1. Branches *1–3* correspond to capillary motions of the liquid, and branches *4*, *5*, *6*, and *7* correspond to aperiodically damped poloidal vortex motions. The general form of the $s=s(W)$ curves for the higher modes of capillary oscillations (with $m>2$) is qualitatively similar. Thus, the roots of Eq. (3.4) form an infinite set with two parameters: the first parameter is the mode number of the capillary oscillations $m$, and the second parameter is the number of the root of Eq. (3.4) for a fixed value of $m$.

It is seen from Fig. 1 that the poloidal vortex motions exhibit aperiodically rapidly damped behavior. The calculations show that their damping decrements increase rapidly as the mode number $m$ increases. As the degree of spheroidal

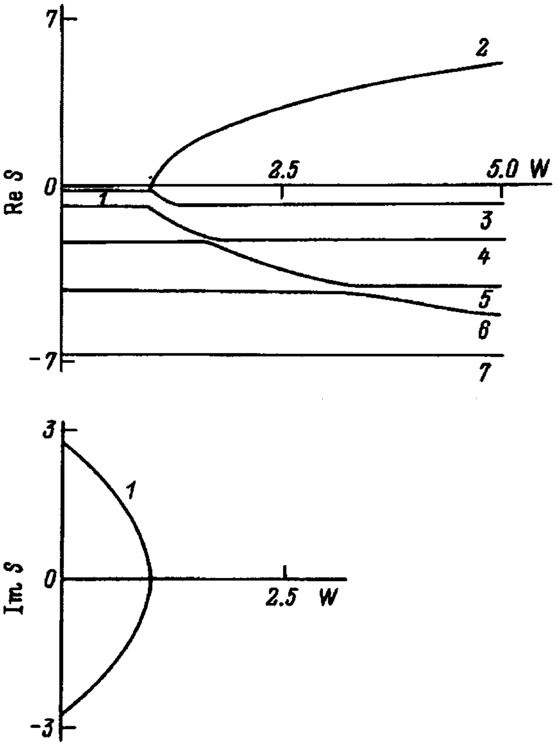

FIG. 1. Dependence of the real and imaginary components of the frequency on the value of the Rayleigh parameter $W$, which characterizes the stability of a droplet against its self-charge.

strain increases (as $e^2$ increases), the decrements of the poloidal motions (as well as the motions corresponding to branches *2* and *3*) decrease slightly (by hundredths of their relative magnitude).

As for the laws governing the realization of the instability of different modes of a highly charged droplet as the degree of spheroidal strain increases, Figs. 2a–c present plots of the dependence of the instability growth rates for the fundamental ($m=2$), sixth ($m=6$), and tenth ($m=10$) modes on the square of the eccentricity $e^2$: $\mathrm{Re}\, s = \mathrm{Re}\, s(e^2)$. The plots presented in Fig. 2 were calculated numerically using (3.4) for $W=16$ [the critical value of the Rayleigh parameter for the realization of droplet instability, i.e., for loss of stability of the fundamental mode, is $W=4$ (Ref. 2)] and for various values of the viscosity of the liquid (curve *1* corresponds to $\nu=0.03$, curve *2* to $\nu=0.1$, curve *3* to $\nu=0.36$, curve *4* to $\nu=0.8$, and curve *5* to $\nu=1.2$).

It is not difficult to see from Fig. 1 that the growth rate of the fundamental mode decreases as $e^2$ increases (generally speaking, this is consistent with Le Chatelier's principle). The growth rates of the higher modes behave differently as $e^2$ increases, depending on the value of the viscosity $\nu$: at small $\nu$ they decrease with increasing $e^2$, and beginning at a certain value of $\nu$, which differs for different modes, they begin to increase with increasing $e^2$. The same behavior is associated with the effect of two different responses of physical parameters of an oscillatory system (which a droplet is) to an increase in $e^2$: on the one hand, an increase in $e^2$ should lead to an increase in the growth rates of the higher modes because of the increase in the surface charge density on the apices of the spheroid,[3] and, on the other hand, the suppressing influence of the viscosity should have the strongest effect on the fast motions of the liquid (caused by the large growth rates of the higher modes). Since the damping decrements of the capillary oscillations decrease with increasing $e^2$, while the growth rates increase, at a certain viscosity these two tendencies should neutralize one another, as was found in the numerical calculations. This effect begins to be manifested for different modes at different values of the viscosity (which we shall mark with an asterisk): for the mode with $m=4$ at $\nu \equiv \nu_{4*} \approx 0.58$, for $m=6$ at $\nu_{6*} \approx 0.165$, and for $m=10$ at $\nu_{10*} \approx 0.068$. The plots of $\mathrm{Re}\, s = \mathrm{Re}\, s(e^2)$ calculated for such values of the parameters are straight lines parallel to the $e^2$ axis, i.e., they do not depend on $e^2$. When $\nu > \nu_{m*}$, the instability growth rate of the $m$th mode increases as $e^2$ increases.

The foregoing statements can be illustrated by an analysis of Eq. (3.4) as $\nu \to 0$. The dispersion relation (3.4) for an ideal liquid ($\nu=0$) takes the form

$$s_0^2 \approx -\left\{ m(m-1)(m+2)\alpha_m - e^2[m^3-(2m-1) \right.$$
$$\left. \times(m+2)|\alpha_m|]\, \frac{m(m+1)}{(2m-1)(2m+3)} \right\}. \qquad (3.5)$$

It is easily seen from (3.5) that the sign of the addition to

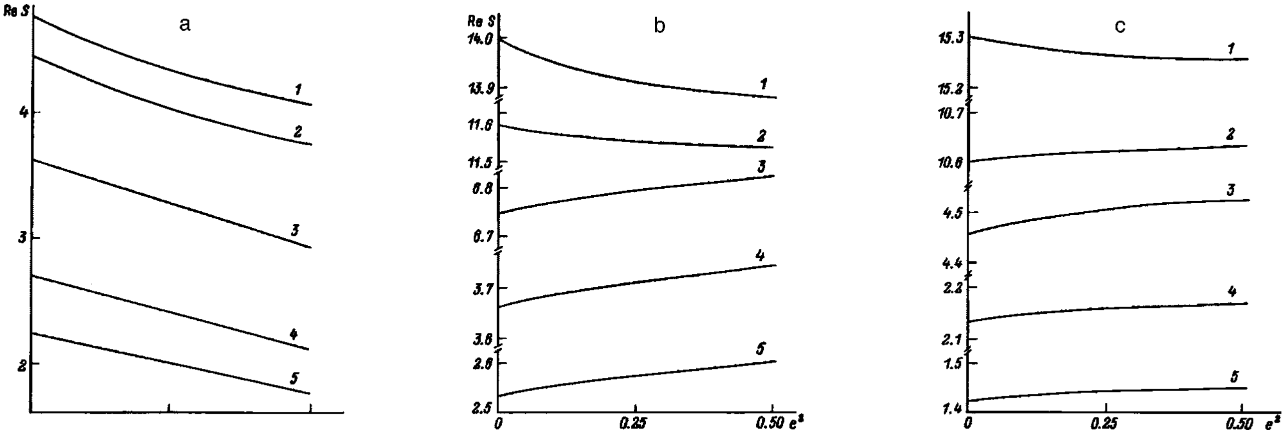

FIG. 2. Dependence of the instability growth rates on the square of the eccentricity.

the instability growth rate associated with the spheroidal shape of the droplet is determined by the sign of the expression in the square brackets. When $m^3>(2m-1)\times(m+2)|\alpha_m|$, as in the case of $W=16$ for $m<4$, the sign of the addition is negative (under this condition the multiplier $|\alpha_m|$ determines the degree of supercriticality of the droplet charge for each of the modes). When the opposite relation holds, i.e., when the mode number is high, the sign of the addition is positive. Similar dependences are also discovered in a numerical analysis of the dispersion relation (3.4) when the viscosity is taken into account (Fig. 2).

It should be noted that the range of variation of $e^2$ from 0 to 0.5, which was used to construct the plots in Fig. 2, was taken only for clarity and for the possibility of qualitatively predicting the possible behavior of the growth rates as the degree of spheroidal strain increases, since Eq. (3.4) was derived for the range $e^2\ll 1$.

The dispersion relation for the oscillatory frequencies of a charged spherical droplet of a viscous liquid[8,11] can be obtained in a relatively simple manner from (3.4) when $e^2=0$.

**4.** Returning now to consideration of the boundary-value problem defined by (2.3) and (2.5) for the scalar function $\Phi_2(\mathbf{r},t)$ and substituting solution (2.8) into condition (2.5), we can easily write an equation of the form

$$D_m^2\left\{\sqrt{\frac{s}{\nu}}\,f_m\left(\sqrt{\frac{s}{\nu}}\right)+(m-1)+e^2\left[-3\sqrt{\frac{s}{\nu}}\,f_m\left(\sqrt{\frac{s}{\nu}}\right)\varkappa_m+\left(\frac{s}{\nu}+(m-1)^2-1+\frac{12}{m(m+1)}\right)\varkappa_m-\frac{2}{3}\right]\right\}=0;$$

$$D_m^2\equiv C_m^2 i_m\left(\sqrt{\frac{s}{\nu}}\right). \qquad (4.1)$$

Condition (4.1) can be satisfied in two ways.

1) By setting all the coefficients $D_m^2$ equal to zero. This corresponds to setting the entire scalar function $\Phi_2(\mathbf{r},t)$ equal to zero. Essentially just this case was considered above when $\Phi_2$ was disregarded. The motions of the liquid appearing in a droplet are then characterized by the dispersion relation (3.4) and do not have a toroidal component, since $U_\varphi=0$.

2) By equating the expression in the curly brackets to zero. If the relation obtained in the latter case is solved for $f_m(\sqrt{s/\nu})$ in the linear approximation with respect to $e^2$, it takes the form

$$\sqrt{\frac{s}{\nu}}\,f_m\left(\sqrt{\frac{s}{\nu}}\right)+(m-1)+e^2\left\{\left[\frac{s}{\nu}+\left((m^2+m-3)+\frac{12}{m(m+1)}\right)\right]\varkappa_m-\frac{2}{3}\right\}=0. \qquad (4.2)$$

Expression (4.2) is the dispersion relation characterizing the spectrum of toroidal vortex motions of the liquid described by $\Phi_2(\mathbf{r},t)$, as well the poloidal motions described by $\Phi_3(\mathbf{r},t)$, due to thermal fluctuations. Figure 3 presents the results of a numerical calculation of the spectrum of toroidal motions with $e^2=0.2$ for the first 22 modes in the form of plots of

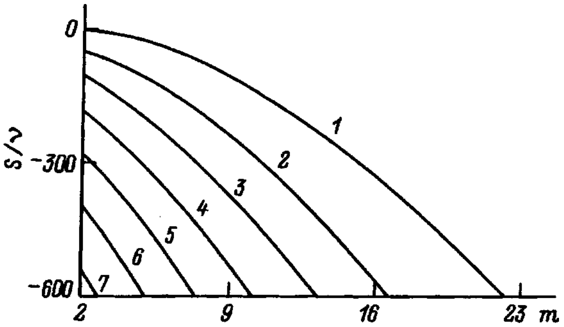

FIG. 3. Dependence of the first several roots of the dispersion relation (4.2) for aperiodically damped toroidal vortex motions of a liquid on the mode number $m$. The numbers of the curves correspond to the roots of the dispersion relation for toroidal vortex motions.

$$\frac{s}{\nu}=\frac{s}{\nu}(m),$$

where $m$ is assumed by convention to vary continuously, rather than discretely, for the first several roots. It is seen that liquid motions of this type are aperiodically strongly damped. The dependence of the damping decrements of the toroidal vortex motions on $e^2$ is very weak and does not exceed hundredths of their relative magnitude in the range of variation of $e^2$ from 0 to 0.5. The set of solutions of the dispersion relation (4.2) has two parameters: the first parameter is the mode number of the capillary oscillations $m$, and the second parameter is the number of the root of Eq. (4.2) for a fixed value of $m$.

It is noteworthy that the boundary-value problem for determining $\Phi_2(\mathbf{r},t)$ is completely autonomous and is not related in any way to the perturbation of the surface $\xi(\theta,t)$ or to $\Phi_1(\mathbf{r},t)$ and $\Phi_3(\mathbf{r},t)$. It follows from this that in the linear formulation of the problem solved here a toroidal motion does not have any influence on the capillary oscillations of the droplet as a whole (i.e., does not have an influence on shaping the surface relief of the droplet) and does not interact with the poloidal motions of the liquid. An interaction between these two types of motions with the formation of a single toroidal-poloidal vortex is realized only in the nonlinear stage.

**5.** According to the results obtained above, the values of the growth rates of the unstable modes with $m>3$ decrease with increasing viscosity and increase as the degree of spheroidal strain increases. This finding points out the possibility of a path for the development of the Rayleigh instability of a highly viscous droplet other than the one described in Ref. 12 for low-viscosity liquids. In the case of low-viscosity liquids, when the stability criterion for the fundamental mode of capillary oscillations is satisfied, the droplet begins to stretch into a spheroid, thereby generating (because of the increase in the local charge density on the apices of the spheroid) the instability of the higher modes of capillary waves. The superposition of the unstable higher modes leads to the formation of emitting tips on the apices of the droplet

(Taylor cones[13,14]), from which the emission of highly disperse, highly charged smaller droplets that carry off charge and mass from the unstable droplet begins. The growth and formation of such emitting tips takes place on a background of the continued (because of the increase in the amplitude of the very unstable fundamental mode) stretching of the droplet into a spheroid, which stops only after the emission of charge and mass begins. In the case of highly viscous liquids, it should be taken into account that, according to the data presented above, the degree of suppression of the instability growth rates of the various modes by the viscosity increases as the mode number increases. Thus, the situation in which the instability growth rates of the higher modes (with $m \geqslant 10$) are smaller than the instability growth rate of the fundamental mode for a highly viscous droplet is real. In this case, before the emitting tips are formed and the emission of charge and mass begins, the droplet is so stretched out that the criterion for its instability toward dividing into halves at high degrees of strain is satisfied.[15,16] As a result, the droplet breaks into two parts of comparable size, each of which is then stable toward its self-charge.

Thus, for highly viscous liquids, in which the dimensionless parameter $\nu[\rho/(R\sigma)]^{1/2}$ is greater than unity, a qualitatively different (in comparison to the case of low-viscosity liquids[12]) path for the realization of the instability of a droplet toward its self-charge is division into two parts of comparable size. It is clear that the fulfillment of the criterion $\nu[\rho/(R\sigma)]^{1/2} \gg 1$ is also possible in the case of low-viscosity liquids for very small droplets. It is not difficult to see that this condition will be satisfied for water droplets when $R \leqslant 0.02\ \mu$m. This phenomenon can occur in, for example, mass-spectrometric experiments on thermally unstable and nonvolatile substances, where highly charged droplets of submicron dimensions that satisfy the condition $\nu[\rho/(R\sigma)]^{1/2} \gg 1$ are emitted from the end of the capillary through which the working liquid is supplied. Consideration of the relaxation of the viscosity[17] enhances the role of the channel for the breakup of a highly charged droplet under discussion,

## APPENDIX

The equilibrium shape of a droplet is determined by the condition of balance between the time-independent pressures acting on the surface, which is obtained from the boundary condition for the normal component of the stress tensor in the absence of motion of the liquid

$$-(p^0 - p^{\mathrm{ex}}) - p_E^0 + p_\sigma^0 = 0.$$

Here $p^0$ is the pressure within the liquid, $p^{\mathrm{ex}}$ is the external pressure, $p_E^0$ is the pressure of the electric forces, and $p_\sigma^0$ is the pressure of the surface-tension forces on the equilibrium droplet surface. It follows from the Navier-Stokes equation that $p^0 = \mathrm{const}$. For an extended spheroid in the linear approximation with respect to the square of the eccentricity of the spheroid $e^2$ it is not difficult to obtain[3]

$$p_E^0 = \frac{Q^2}{8\pi}\left[1 + \frac{1}{3}e^2(3\cos^2\theta - 1)\right];$$

$$p_\sigma^0 = 2\left[1 + \frac{1}{3}e^2(3\cos^2\theta - 1)\right].$$

To maintain the equilibrium spheroidal shape of a droplet, an external pressure $p^{\mathrm{ex}}$ which does not depend on the time or the shape of the droplet surface must be created. This pressure is specified by the pressure balance condition

$$p^{\mathrm{ex}} = p^0 - 2\left(1 - \frac{Q^2}{16\pi}\right)\left[1 + \frac{1}{3}e^2(3\cos^2\theta - 1)\right].$$

This pressure can be, for example, an ultrasonic pressure.

[1] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk Mekh. Zhidk. Gaza, No. 3, 3 (1994).

[2] J. W. Rayleigh, Philos. Mag. **14**, 184 (1882).

[3] A. I. Grigor'ev, Zh. Tekh. Fiz. **55**, 1273 (1985) [Sov. Phys. Tech. Phys. **30**, 736 (1985)].

[4] K. J. Cheng, Phys. Lett. A **112**, 392 (1985).

[5] S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **22**(4), 84 (1996) [Tech. Phys. Lett. **22**(2), 171 (1996)].

[6] S. O. Shiryaeva and A. I. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **21**(16), 17 (1995) [Tech. Phys. Lett. **21**(8), 639 (1995)].

[7] S. O. Shiryaeva and I. D. Grigor'eva, Pis'ma Zh. Tekh. Fiz. **20**(6), 1 (1994) [Tech. Phys. Lett. **20**(3), 214 (1994)].

[8] S. O. Shiryaeva, A. É. Lazaryants *et al.*, IMRAN Preprint No. 27 [in Russian], Institute of Microelectronics, Russian Academy of Sciences, Yaroslavl (1994), 126 pp.

[9] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum: Irreducible Tensors, Spherical Harmonics, Vector Coupling Coefficients, 3nj Symbols*, World Scientific, Singapore-Teaneck, NJ (1988).

[10] S. O. Shiryaeva A. I. Grigor'ev *et al.*, IMRAN Preprint No. 29 [in Russian], Institute of Microelectronics, Russian Academy of Sciences, Yaroslavl (1995), 344 pp.

[11] A. I. Grigor'ev and A. É. Lazaryants, Izv. Akad. Nauk SSSR Mekh. Zhidk. Gaza, No. 5, 11 (1991).

[12] A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **61**(3), 19 (1991) [Sov. Phys. Tech. Phys. **36**, 258 (1991)].

[13] A. I. Grigor'ev and S. O. Shiryaeva, Inzh. Fiz. Zh. **61**, 632 (1991).

[14] S. I. Shevchenko, Zh. Tekh. Fiz. **60**(2), 54 (1990) [Sov. Phys. Tech. Phys. **35**, 167 (1990)].

[15] A. I. Grigor'ev, S. O. Shiryaeva, and I. D. Grigor'eva, Zh. Tekh. Fiz. **61**(4), 25 (1991) [Sov. Phys. Tech. Phys. **36**, 385 (1991)].

[16] S. O. Shiryaeva, V. A. Koromyslov, and A. I. Grigor'ev, Elektron. Obrab. Mater., No. 5, 55 (1995).

[17] S. O. Shiryaeva and O. A. Grigor'eva, Pis'ma Zh. Tekh. Fiz. **21**(9), 67 (1995) [Tech. Phys. Lett. **21**(5), 346 (1995)].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Ion current density in an expanding laser plasma

S. V. Latyshev and Yu. N. Cheblukov

*Institute of Theoretical and Experimental Physics, 117259 Moscow, Russia*

(Submitted November 11, 1996)
Zh. Tekh. Fiz. **68**, 28–32 (April 1998)

The dependence of the ion current density in an expanding laser plasma on the parameters of the laser radiation, the target material, and the distance from the target is investigated theoretically. Calculations needed to design laser ion sources for accelerator technology are performed. An explanation for the two-peaked shape of the collector signals observed in some experiments with a laser plasma is given. Additional possibilities for obtaining information on the experimental conditions from collector signals are considered. © *1998 American Institute of Physics.* [S1063-7842(98)00604-7]

In this paper we calculate the ion current density $j = Zn_1u$ in a laser plasma expanding along a normal to the target surface and investigate the dependence of this quantity on the parameters of the laser radiation, the target material, and the distance from the target ($Z$ is the mean charge of the ions, $n_1$ is the ion density, and $u$ is the expansion velocity of the ions along the normal).

There have been only a few calculations modeling the collector signals in experiments with a laser plasma,[1,2] and they did not address the problem of elucidating the dependence of the ion currents on the parameters of the experiment. This topic is important both for designing laser ion sources for accelerator technology and for interpreting the results of collector diagnostics of expanding laser plasmas.

## QUASI-TWO-DIMENSIONAL MODEL OF A LASER PLASMA

All the calculations in the present work were performed using the quasi-two-dimensional two-temperature hydrodynamic model of a laser plasma proposed in Ref. 3. This model was previously used with success to calculate the charge composition of ions in an expanding laser plasma,[4] the temperature of the laser plasma,[5] and several recombination effects.[6,7] According to this model, a laser plasma is simulated by ''large particles'' of equal mass in the form of thin disks. Under the action of the forces of the gas-kinetic pressure, the disks can move along a normal to the target, and their radius can vary. The transverse expansion of the disks is considered in the self-similar approximation, i.e., the radial expansion velocities of inner points on a disk are related to the radial velocity of the disk boundary by the law $v_r(\xi) = u_r\xi/r$, where $\xi$ is the radial coordinate of the inner point on the disk, and $r$ and $u_r$ are the radius and radial velocity of the disk boundary. Thus, the quasi-two-dimensional model is something midway between the pure one-dimensional model of a plasma and the cylindrically symmetric two-dimensional model.

The model incorporates the following physical processes: the absorption of laser radiation by an inverse bremsstrahlung mechanism and reflection from a plasma layer with the critical density;[8] electronic heat conduction with consideration of the classical restriction of the heat flux;[9] heat exchange between electrons and ions in elastic collisions;[10] ionization processes in the mean-charge approximation with consideration of electron-impact ionization, three-body recombination, dielectronic recombination, and photorecombination;[11] and the energy balance for the inelastic processes. In the energy balance for the inelastic processes the energy returned to plasma electrons upon the three-body recombination of ions through highly excited states is taken into account using the formula proposed in Ref. 12,

$$E_t = 7\times10^{-22}(n_e/Z)^{2/3}/\Theta_e[\mathrm{erg}],$$

where $\Theta_e$ is the electron temperature in electron-volts.

This is the essential difference between the present model and the models of other authors.

## DEPENDENCE OF THE ION CURRENT DENSITY ON THE PARAMETERS OF THE LASER RADIATION, THE TARGET MATERIAL, AND THE DISTANCE FROM THE TARGET

The simplest functional dependence of the ion current density $j = eZn_1u$ on the parameters of the experiment can be obtained by assuming that $n_1 \sim W_{\mathrm{las}}/(ZT_eL^3)$ and $u \sim (ZT_e/m_i)^{1/2}$, where $W_{\mathrm{las}}/ZT_e$ is an estimate of the number of ions generated by the laser pulse, $L$ the distance from the target, and the expansion velocity of the ions is estimated from the velocity of sound $u_s$ in the heating phase. Hence we obtain

$$j \sim eW_{\mathrm{las}}L^{-3}m_i^{-1/2}(Z/T_e)^{1/2}. \qquad (1)$$

It is seen from formula (1) that the ion current density is proportional to the laser pulse energy $W_{\mathrm{las}}$, is inversely proportional to $m_i^{1/2}$, and depends most strongly ($\sim L^{-3}$) on the distance to the target. All the remaining dependences are less significant.

Let us consider the main dependences of the ion current density on the parameters of the heating radiation in the example of an expanding gold plasma heated by a rectangular pulse from a $CO_2$ laser. All the calculations of the ion current

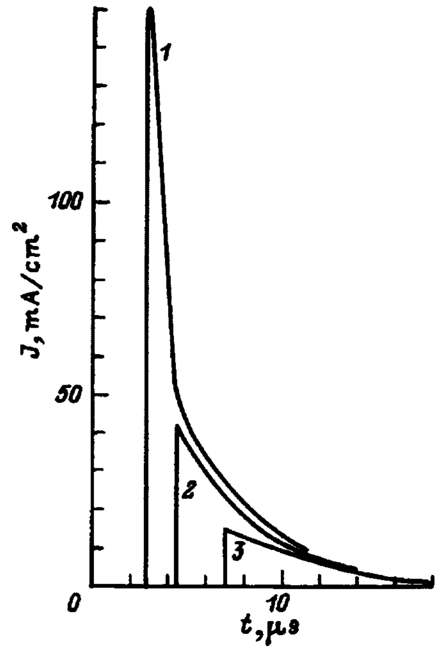

FIG. 1. Dependence of the ion pulse on the power of the laser radiation. Au, $d=1$ mm, $\tau=100$ ns, $\lambda=10.6$ $\mu$m, $L=1$ m; $P$, W/cm$^2$: *1*—$10^9$, *2*—$3\times10^8$, *3*—$10^8$.

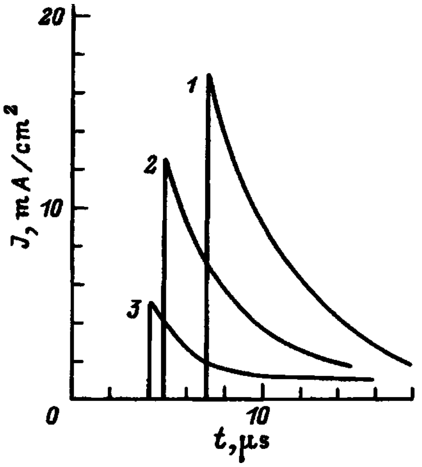

FIG. 2. Dependence of the ion pulse on the diameter of the focal spot of the laser radiation. Au, $P=10^8$ W/cm$^2$, $\tau=100$ ns, $\lambda=10.6$ $\mu$m, $L=1$ m; $d$, mm: *1*—1.0, *2*—0.6, *3*—0.2.

densities presented in this section correspond to a distance from the target equal to 1 m, unless stated otherwise. In addition, for brevity, the time dependence of the ion current density will be called the ion pulse below.

Figure 1 presents the ion pulses for various laser powers $P$ in the range $10^8-10^9$ W. The laser pulse duration and the focal spot diameter were $\tau=100$ ns and $d=1$ mm. The calculations presented in Fig. 1 attest to an increase in the amplitude of the ion pulse in accordance with estimate (1) and to a decrease in the delay time of the beginning of the ion pulse as the power of the laser radiation increases (the delay time is measured from the beginning of the laser pulse). Such dependences are attributed to increases in the laser pulse energy and, consequently, in the number of ions generated by the laser radiation, as well as to increases in the plasma temperature and, consequently, in the expansion velocity of the ions.

Figure 2 presents the ion pulses for various diameters of the focal spot of the laser radiation and equal values of the other parameters. The results presented in Fig. 2 attest to a decrease in the amplitude of the ion pulse and the delay time of the ion pulse as the focal spot diameter decreases. There is also a fairly simple explanation for these dependences. As the focal spot diameter decreases, the plasma temperature and, consequently, the expansion velocity of the ions increase, causing a decrease in the delay time of the ion pulse. On the other hand, when the focusing is sharper and the energy of the laser radiation is fixed, the amount of plasma produced in the heating phase decreases, and, more importantly, the lateral expansion of the plasma increases. The latter leads to a considerable decrease in the number of ions traveling along a normal to the target in the far zone, which is the decisive factor causing a decrease in the amplitude of the ion pulse as opposed to the increase in it due to the increase in the expansion velocity.

The dependence of the ion pulse on the atomic weight of the target is presented in Fig. 3. As the atomic weight of the target increases, the amplitude of the ion pulse decreases in accordance with estimate (1), and the delay time of the ion pulse increases. These dependences are attributed to the decreases in the expansion velocity and the number of ions as the atomic weight of the target increases.

The dependence of the ion pulse on the wavelength of the laser radiation is presented in Fig. 4. It is the weakest of all the dependences considered. The relatively small increase in the amplitude and delay time of the ion pulse with decreasing wavelength is attributed to some decrease in the temperature and lateral expansion of the plasma as the wavelength of the laser radiation is diminished. When the focusing is sharper, the decrease in the quantity of material vaporized due to the decrease in the absorption of long-wavelength laser radiation in comparison to short-wavelength radiation will act in the same direction.

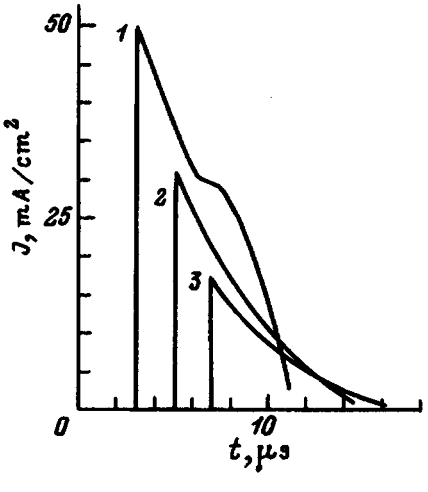

FIG. 3. Dependence of the ion pulse on the target material. $P=10^8$ W/cm$^2$, $d=1$ mm, $\tau=100$ ns, $\lambda=10.6$ $\mu$m, $L=1$ m; *1*—C, *2*—Fe, *3*—Au.

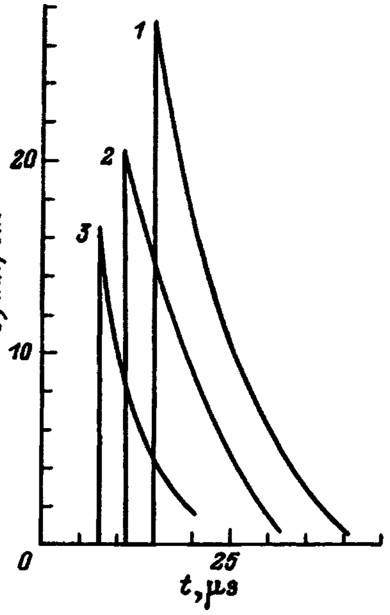

FIG. 4. Dependence of the ion pulse on the wavelength of the laser radiation. Au, $P=10^8$ W/cm², $\tau=100$ ns, $L=1$ m; λ, μm: *1*—1.06, *2*—3.0, *3*—10.6.

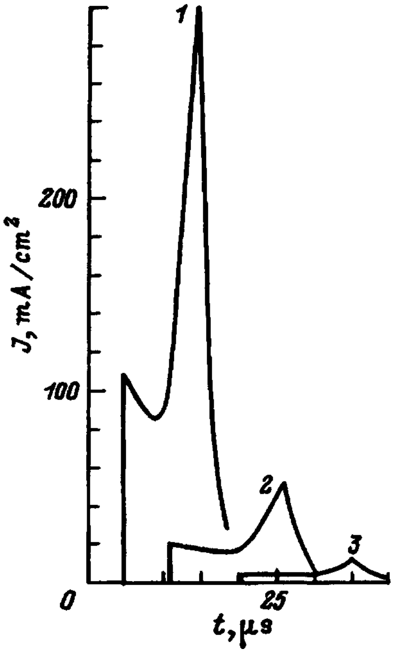

FIG. 6. Dependence of the ion pulse on the distance from the target. Au, $P=10^8$ W/cm², $d=1$ mm, $\tau=1000$ ns; $L$, m: *1*—1, *2*—2, *3*—3.

The dependence of the ion pulse on the laser pulse duration turned out to be highly significant. The results of the calculation of the ion pulses for laser pulse durations from 100 to 1000 ns in Fig. 5 show that in the general case the ion pulse consists of two peaks: a fast peak and a slow peak. At small laser pulse durations the ion pulse consists of one fast peak. As the laser pulse duration increases, a second peak appears at slow velocities. When the pulse duration is sufficiently long, the slow peak begins to strongly predominate over the fast peak. The fast peak is caused by ions whose origin is associated with the initial nonstationary heating regime of the laser plasma, and the slow peak is associated with ions which appear already in the steady stationary heating regime. The appearance of the slow peak requires the laser pulse duration to be much greater than the characteristic hydrodynamic time, i.e., $\tau_{\text{las}} > 3-5d/u_s$.

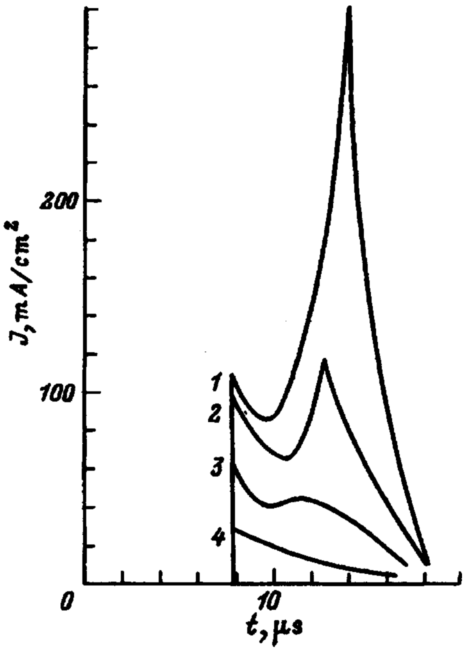

FIG. 5. Dependence of the ion pulse on the duration of the laser radiation. $P=10^8$ W/cm², $d=1$ mm, $\lambda=10.6$ μm, $L=1$ m; τ, ns: *1*—1000, *2*—600, *3*—300, *4*—100.

Single-peaked and two-peaked collector signals have been recorded in different laser plasma experiments employing collector diagnostics.[13,14] In addition, multiple-peaked collector signals have been observed in several experiments, apparently due to the complex form of the laser pulse and to the electrostatic mechanism of ion acceleration at high laser radiation flux densities.[15]

The ion pulse depends most strongly on the distance to the target. The corresponding calculations presented in Fig. 6 show that as the distance to the target increases, there are sharp decreases in the amplitudes of both the fast and slow peaks in the ion pulse approximately according to a $j \sim L^{-3}$ law, as well as a linear increase in the delay time of the ion pulse.

## INTERPRETATION OF THE COLLECTOR DIAGNOSTICS OF THE LASER PLASMA

One clear and very important conclusion that can be drawn from the results considered in the preceding section is that the ion current density is very sensitive to most of the parameters of the experiment, such as the distance from the target, the power and duration of the laser radiation, the target material, etc. This suggests that additional information regarding the experimental conditions can be obtained by achieving the closest possible fit between the calculated data and the experiment. However, any experiment whose conditions are known inexactly can be ‘‘reconciled’’ fairly easily

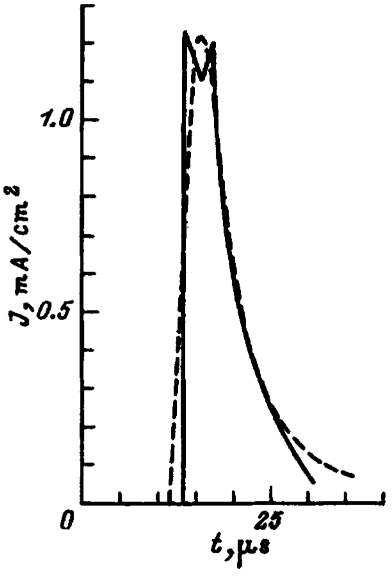

FIG. 7. Comparison of experimental and theoretical current density pulses of carbon ions. $P=5\times10^7$ W/cm², $\tau=35$ ns, $d=0.3$ mm, $\lambda=1.06$ μm, and $L=2$. Solid curve—theory, dashed line—experiment.

with calculated results by correcting the calculated parameters in a particular direction. We shall demonstrate this situation in the case of two experiments.[13,14]

In Ref. 13 a neodymium laser with a maximum pulse energy $W\sim30$ J, a laser pulse duration at half maximum $\tau\sim25$ ns, and a focal spot diameter $d\sim300$ μm was used. The most abundant information is available for the collector signals of carbon ions at a distance of 2 m from the target, which were obtained in experiments in which the energy of the laser radiation was reduced by a factor of 5 using a calibrated absorber. In addition, according to the authors' estimates, an additional twofold decrease in the laser pulse energy occurred in the optical system due to the geometric divergence of the radiation. The radiation was reduced by about two more fold as a result of reflection of the laser radiation from the numerous elements of the optical system. Thus, the energy of the radiation observed by the target was apparently equal to 1–2 J. A set of calculations showed that the calculation in which the laser pulse energy was 1.1 J, the focal spot diameter was 300 μm, and the laser pulse was modeled by a half sinusoid with a duration along the base of 35 ns agrees most closely with the experiment. This calculation of the collector signal and the corresponding experimental signal are presented in Fig. 7.

A $CO_2$ laser, whose pulse consisted of a peak of duration 200–300 ns with an energy of about 10 J and a tail of duration 2 μs with an energy up to 20 J, was used in Ref. 14. The collector signal for zirconium ions at a distance of 2 m from the target had a two-peaked structure. The corresponding set of calculations shows that the theoretical collector signal most closely approximates the experimental signal when the calculation parameters have the following values: a laser pulse peak approximated by a half sinusoid with a duration along the base equal to 250 ns and an energy of 6.4 J, a tail approximated by a straight line descending from the half height of the peak to zero with a duration of 1.75 μs and an energy of 18 J, and a focal spot diameter equal to 400 μm. The corresponding experimental and calculated collector signals are presented in Fig. 8. The calculations showed that the fast peak of the collector signal is caused by ions which formed under the action of the laser radiation peak and that the slow peak of the collector signal is caused by ions from the tail of the laser pulse.

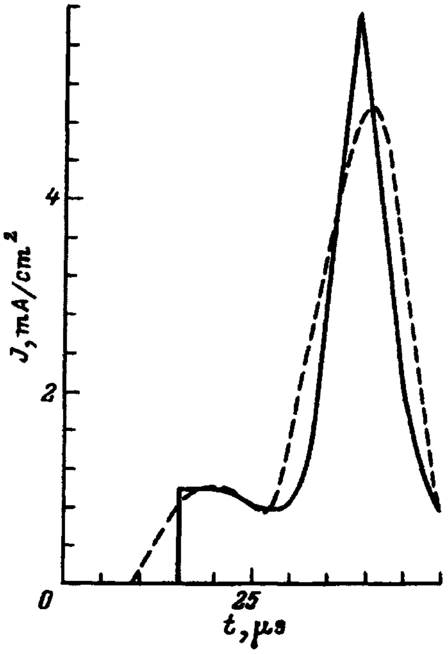

FIG. 8. Comparison of experimental and theoretical current density pulses of zirconium ions. $P=4\times10^7$ W/cm², $\tau=2000$ ns, $d=0.4$ mm, $\lambda=10.6$ μm, and $L=2$ m. Solid curve—theory, dashed line—experiment.

Calculations of both experiments showed that the experimental and calculated collector signals can be closely reconciled without departing from the reasonable ranges of the calculated parameters. The set of optimal parameters was quite rigidly specified in both cases. For example, twofold variation of any of the parameters of the laser radiation, such as the energy, the duration, and the focal spot diameter led to a failure to observe satisfactory agreement between theory and experiment for any variations of the other parameters.

## CONCLUSIONS

The quasi-two-dimensional hydrodynamic model of a laser plasma has been found to be suitable for calculating the ion current density in an expanding laser plasma despite the very rough simulation of the lateral spreading of the plasma.

The dependence of the ion current density in an expanding laser plasma on the parameters of the laser radiation, the target material, and the distance from the target has been investigated theoretically. Calculations needed to design laser ion sources for accelerator technology have been performed. An explanation for the two-peaked shape of the collector signals observed in some experiments with a laser plasma has been given.

The possibility of extracting additional information on the experimental conditions from collector signals using a series of approximation calculations has been demonstrated.

[1] Yu. N. Erema, S. V. Latyshev, V. V. Petrov *et al.*, ITÉF Preprint No. 199 [in Russian], Institute of Theoretical and Experimental Physics, Moscow (1987), 16 pp.

[2] K. N. Makarov, Yu. A. Satov, A. P. Strel'tsov *et al.*, Zh. Éksp. Teor. Fiz. **106**, 1649 (1994) [JETP **79**, 891 (1994)].

[3] S. V. Latyshev, ITÉF Preprint No. 66 [in Russian], Institute of Theoretical and Experimental Physics, Moscow (1983), 20 pp.

[4] A. A. Golubev, S. V. Latyshev, and B. Yu. Sharkov, Kvantovaya Elektron. **11**, 1854 (1984) [Sov. J. Quantum Electron. **14**, 1242 (1984)].

[5] S. V. Latyshev and I. V. Rudskoĭ, ITÉF Preprint No. 2 [in Russian], Institute of Theoretical and Experimental Physics, Moscow (1986), 16 pp.

[6] S. V. Bobashev, S. V. Latyshev, I. V. Rudskoĭ *et al.*, Fiz. Plazmy **13**, 1383 (1987) [Sov. J. Plasma Phys. **13**, 801 (1987)].

[7] Yu. N. Erema, S. V. Latyshev, and A. V. Shumshurov, ITÉF Preprint No. 28 [in Russian], Institute of Theoretical and Experimental Physics, Moscow (1988), 4 pp.

[8] Yu. V. Afanas'ev and O. N. Krokhin, Tr. Fiz. Inst. Akad. Nauk SSSR **52**, 118 (1970).

[9] L. Spitzer Jr., *Physics of Fully Ionized Gases*, 2nd ed., Interscience, New York (1962) [Russian trans., Mir, Moscow (1965), 212 pp.].

[10] V. E. Golant, A. P. Zhilinskii, and S. A. Sakharov, *Fundamentals of Plasma Physics*, Wiley, New York (1980) [Russian orig., Atomizdat, Moscow, 1977, 384 pp.].

[11] I. I. Sobelman, L. A. Vainshtein, and E. A. Yukov, *Excitation of Atoms and Broadening of Spectral Lines*, Springer-Verlag, Berlin (1981) [Russian orig., Nauka, Moscow (1979), 319 pp.].

[12] S. V. Latyshev and I. V. Rudskoĭ, Fiz. Plazmy **11**, 1175 (1985) [Sov. J. Plasma Phys. **11**, 669 (1985)].

[13] O. B. Anan'in, A. M. Baldin, Yu. A. Bykovskiĭ *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **19**, 19 (1974) [JETP Lett. **19**, 10 (1974)].

[14] A. A. Golubev, Yu. N. Erema, B. Yu. Sharkov, and A. V. Shumshurov, ITÉF Preprint No. 134 [in Russian], Institute of Theoretical and Experimental Physics, Moscow (1988), 20 pp.

[15] A. W. Ehler, J. Appl. Phys. **46**, 2464 (1975).

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Features of the development of pulsed microwave discharges in various gases in a quasioptical beam

L. P. Grachev, I. I. Esakov, and K. V. Khodataev

*Moscow Institute of Radio Engineering, Russian Academy of Sciences, 113519 Moscow, Russia*

(Submitted January 8, 1997)
Zh. Tekh. Fiz. **68**, 33–36 (April 1998)

Photographs of pulsed microwave discharges initiated by a metallic sphere placed at the focus of a quasioptical electromagnetic beam with linear polarization of the field in air, sulfur hexafluoride, hydrogen, and helium under a pressure of several hundred Torr are presented. The common and distinctive features of the discharges in these gases are noted. © *1998 American Institute of Physics.* [S1063-7842(98)00704-1]

## INTRODUCTION

As a pulsed gas microwave discharge in a focused microwave TEM wave develops, it can pass through such successive stages: as the diffuse, ionization-overheating, streamer, resonant, and cumulative stages.[1] At a fixed discharge duration $t_d$, their sequence remains unchanged, but the number of stages realized increases as the gas pressure $p$ and the amplitude of the electric component of the initial microwave field $E_0$ are increased. The last of the stages realized at the time of completion of the microwave pulse largely determines the outward appearance of the plasmoid that develops. In this sense the terms diffuse, ionization-overheating, streamer, etc. can be used to describe the form of the discharge.[2]

The main physical factors which determine the properties of a discharge in each of the stages just enumerated do not depend on the kind of gas. However, while the general picture of the development of discharges in different gases is qualitatively consistent, its specific details may differ.[3,4] The distinctive features derive, for example, from quantitative differences in the level of the breakdown field $E_{\mathrm{bd}}$, in the values of the kinetic coefficients, in the presence or absence of ionizing ultraviolet radiation from the discharge region, etc.

The experiments in Ref. 1 and 5 showed that, in principle, there are no significant differences between electrodeless and initiated discharges. At the same time, the employment of an initiator, especially at high $p$, permits significant expansion of the range of variation of $E_0$. Moreover, the use of an initiator in the form of a metallic sphere, for example, permits effective calibration of the absolute level of the field according to the method described in Ref. 6.

To supplement the data on microwave discharges in various gases, in this paper we present photographs of microwave discharges initiated by a sphere in a focused traveling electromagnetic wave in air, sulfur hexafluoride ($SF_6$), hydrogen ($H_2$), and helium (He) at values of $p$ equal to hundreds of Torr. Their common features and differences are noted. Among the gases just listed, air and $SF_6$ are electronegative, and the electron attachment coefficient is significantly greater in $SF_6$ than in air. The composition of the molecules in air and $SF_6$ is more complicated than in $H_2$, and He is an atomic gas. Thus, the spectrum of differences between the gases investigated is fairly broad.

## EXPERIMENTAL CONDITIONS

Pulsed TEM radiation having a rectangular envelope with linear polarization and a wavelength of 8.5 cm was used in the experiments.[5] In the focal region the field had an approximately Gaussian distribution with a characteristic scale equal to several centimeters in the plane perpendicular to the Poynting vector $\mathbf{\Pi}$. Along $\mathbf{\Pi}$ it was uniform and also had a scale of several centimeters. The maximum value of $E_0$ was about 6.5 kV/cm. The microwave pulse had a duration of 40 $\mu$s. The pauses between pulses lasted at least a minute.

A hermetically sealed quartz tube with flat endcaps made from optically transparent glass was positioned symmetrically to the axis of the beam at its focus. The tube had an inner diameter of 8 cm, a wall thickness of 0.4 cm, and a length of 50 cm. Its axis was perpendicular to $\mathbf{\Pi}$ and $\mathbf{E}_0$. The tube, which was preliminarily evacuated to $10^{-1}$ Torr, was filled with the gas under investigation, whose pressure $p$ was fixed to within $\pm 0.75$ Torr. The tube was surrounded by air at atmospheric pressure.

A lead sphere with a diameter of 0.25 cm on a thread with a thickness of $10^{-2}$ cm was placed in the tube at the focus of the electromagnetic beam. The thread extended along a diagonal of the central cross section of the tube perpendicular to $\mathbf{E}_0$. The surface of the sphere was the source of initiating electrons. For this purpose it was illuminated by ultraviolet (UV) radiation with a duration of 10 $\mu$s. In the experiments the time of breakdown of the gases investigated coincided with the beginning of the UV pulse. By shifting it, we could vary the discharge time $t_d$, which extends from the onset of the pulse of UV illumination to the end of the microwave pulse.

The values of $E_0$ and $t_d$ were adjusted so that the discharge, which began on the sphere, would not reach the inner surface of the tube.

The discharge was photographed through the endcaps of the tube. The exposure time exceeded $t_d$. In the photographs

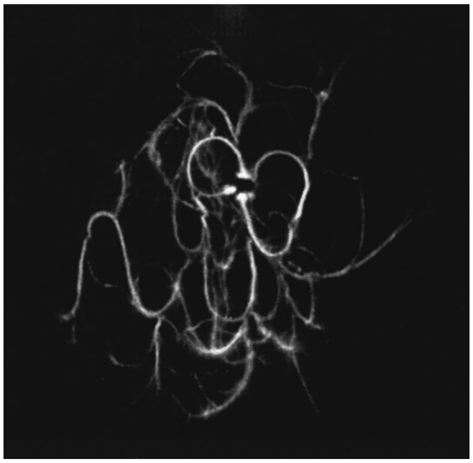

FIG. 1. Microwave discharge in air at $p=300$ Torr with $E_0=6.5$ kV/cm.

presented below $\mathbf{E}_0$ is vertical, and $\mathbf{\Pi}$ is directed from right to left. The diameter of the sphere serves as a geometric scale on them. On all the images $t_p=10\ \mu$s.

## OBSERVATION RESULTS

Figure 1 presents a discharge in air at 300 Torr with $E_0=6.5$ kV/cm. It is seen that it is a typical streamer burning in the microwave range.[5] The individual portions of the main streamers and their ends are ''feathered'' by thin short plasma channels with a weak diffuse corona.

At this pressure air did not undergo breakdown at $E_0<5.7$ kV/cm (the discharge for 300 Torr and $E_0=5.7$ kV/cm$\simeq E_{\text{bd}}$ is presented in Fig. 2). Therefore, in the experiment $E_{\text{bd}}\simeq 1/3E_k$, where $E_k=40p$ is the so-called critical breakdown field (here and in the analogous formulas below, $p$ is expressed in Torr and $E_k$ in V/cm). We recall that at the poles of the sphere, where $E_0$ is perpendicular to its surface, the field equals $3E_0$. The measured value of $E_{\text{bd}}$ is only 17% higher than the value following from the theory.[6] Thus, the presence of the tube at the focus of the electromagnetic beam scarcely distorted its structure or the strength of the original field.

FIG. 2. Microwave discharge in air at $p=300$ Torr with $E_0=5.7$ kV/cm.

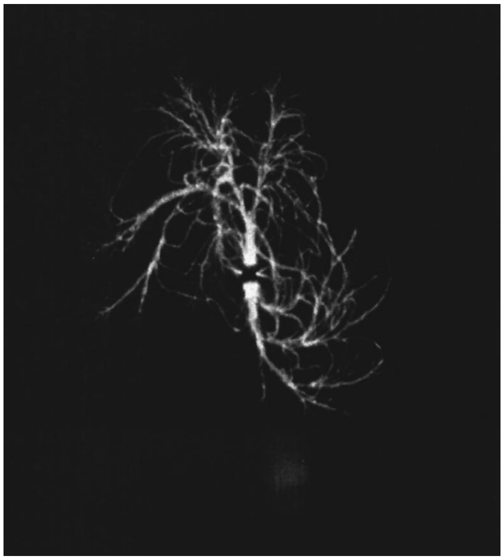

FIG. 3. Microwave discharge in sulfur hexafluoride at $p=90$ Torr with $E_0=5$ kV/cm.

Figure 3 shows a discharge in $SF_6$ at 90 Torr with $E_0=5$ kV/cm. It is seen that it is also a streamer discharge. The feathering channels are longer than those in the air discharges. This creates the impression of a greater volume density of discharge channels. Their diffuse ''fur'' is less pronounced than in air. The presence of a large number of comparatively long feathering channels is the main distinguishing feature of the $SF_6$ discharges.

When $E_0=5$ kV/cm, $SF_6$ undergoes breakdown only at $p\leqslant 105$ Torr. This corresponds to $E_k=150p$. This value is approximately 30% greater than the values of $E_k$ for sulfur hexafluoride usually presented in handbooks. For example, Ref. 7 gives a value $E_k=100p$, and Ref. 8 gives $117p$.

Figure 4 presents a discharge in $H_2$ at 300 Torr with $E_0=5$ kV/cm. Under these conditions the hydrogen discharge is also of the streamer type. The discharge channels exhibit practically no streamer feathering or diffuse fur, creating the impression that they are rarefied. The absence of the fine details in the $H_2$ discharge is its principal outward distinction.

When $p=300$ Torr, $H_2$ undergoes breakdown only at $E_0\geqslant 4.4$ kV/cm, i.e., $(E_0/p)_{\text{bd}}\simeq 15p$, and at $p=660$ Torr, it undergoes breakdown only at 6.5 kV/cm, which gives $(E_0/p)_{\text{bd}}\simeq 10p$. The latter value coincides with the value of $(E_0/p)_{\text{bd}}$ in a uniform field presented, for example, in Refs. 8 and 9. It can be assumed that in the experimental setup under consideration the function of the sphere is confined only to ensuring the presence of breakdown-initiating electrons and that the region of the tripled field on its poles is simply not ''noticed'' by electrons, which readily diffuse in $H_2$. In addition, in hydrogen, even at a value of $p$ amounting to hundreds of Torr, inhomogeneity of the microwave beam with a characteristic scale equal to several centimeters must be taken into account in the calculation of the breakdown

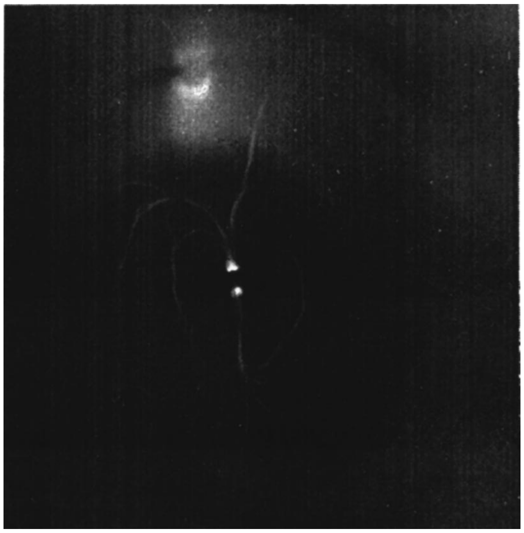

FIG. 4. Microwave discharge in hydrogen at $p = 300$ Torr with $E_0 = 5$ kV/cm.

fields. The experimentally detected increase in $(E_0/p)_{\mathrm{bd}}$ with decreasing $p$ is attributed to just such inhomogeneity.

Finally, Fig. 5 presents a discharge in He at 300 Torr with $E_0 = 4.4$ kV/cm. It is seen that under these conditions the He discharge is realized practically in the diffuse form. When $E_0$ is increased to $E_{\max}$ at this value of $p$, the discharge takes the ionization-overheated form and still does not become a streamer discharge. This finding supports the conclusions of Ref. 4, in which the realization of a streamer discharge in He was indicated only at $p \geqslant 500$ Torr. In fact, Fig. 6 shows a discharge in He at 760 Torr with $E_0 = 5.5$ kV/cm. The figure reveals that it is already a streamer discharge, although it has a significant diffuse background, on which there are only outlines of feathering channels.

Under the conditions of the experiment at $p = 300$ Torr, He undergoes breakdown at $E_0 \geqslant 4.4$ V/cm, i.e., $(E_0/p)_{\mathrm{bd}} = 15$, and in the experiment at 760 Torr it undergoes breakdown at 5.5 kV/cm, i.e., $(E_0/p)_{\mathrm{bd}} = 7$. According to various published sources, $(E_0/p)_{\mathrm{bd}} = 3$ (Ref. 8), 4.7 (Ref. 4), and 1–3 (Ref. 9). Therefore, in He, just as in $H_2$, electron diffusion plays a significant role in the evaluation of the breakdown fields under spatially inhomogeneous conditions.

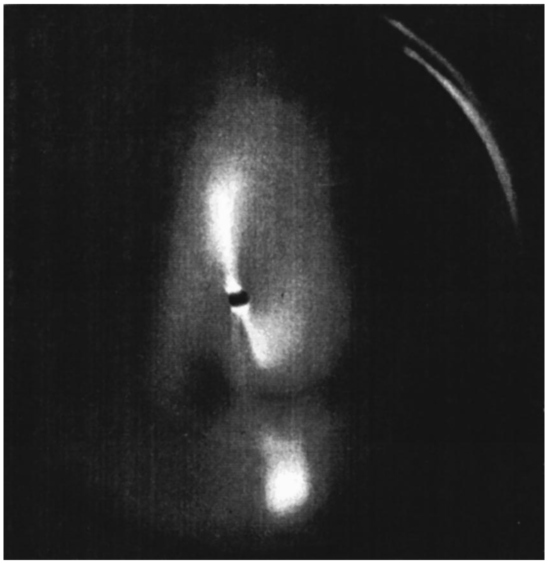

FIG. 5. Microwave discharge in helium at $p = 300$ Torr with $E_0 = 4.4$ kV/cm.

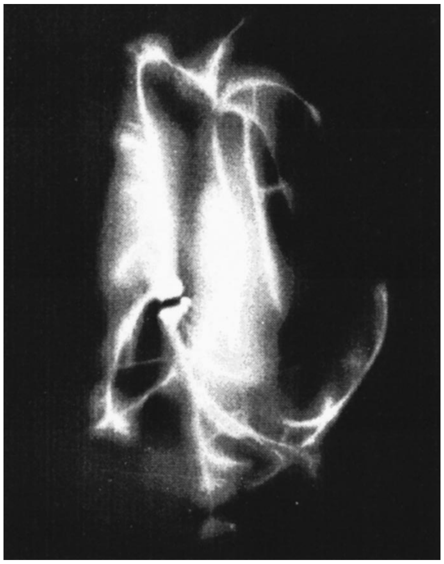

FIG. 6. Microwave discharge in helium at $p = 760$ Torr with $E_0 = 5.5$ kV/cm.

## CONCLUSIONS

Thus, the main features microwave discharges in air, $SF_6$, $H_2$, and He at pressures equal to hundreds of Torr in a focused TEM beam are similar. At the same time, the discharges in the electronegative gases (air and $SF_6$) take the streamer form already at relatively low values of $p$ (less than 100 Torr), but a discharge in He, for example, takes this form only at $p > 500$ Torr. A significant difference between the streamer forms of discharges in different gases is the presence or absence in them of a comparatively small-scale streamer structure, which is manifested as feathering of the main discharge channels. For example, while it is practically absent in $H_2$, the rudiments of it can already be seen in He. The small-scale streamer structure is more pronounced in air, and in $SF_6$ it largely determines the entire outward appearance of the discharge. In He at values of $p$ close to atmospheric pressure, the diffuse background and the diffuse corona around the streamer channels are still significant, while in $H_2$, for example, they are already absent at $p \simeq 300$ Torr.

The experiment showed that $E_k = 150p[\mathrm{Torr}]$ V/cm in $SF_6$. This value is higher than the value given in handbooks.

The values of $(E_0/p)_{\mathrm{bd}}$ in $H_2$ and He obtained in the experiments can serve as reference points in devising a theory for the microwave breakdown of these gases with consideration of the diffusive phenomena in spatially inhomogeneous Gaussian beams.

[1] L. P. Grachev, I. I. Esakov, G. I. Mishin, and K. V. Khodataev, Zh. Tekh. Fiz. **66**(7), 32 (1996) [Tech. Phys. **41**, 652 (1996)].
[2] L. P. Grachev, I. I. Esakov, G. I. Mishin, and A. B. Fedotov, Zh. Tekh. Fiz. **59**(10), 149 (1989) [Sov. Phys. Tech. Phys. **34**, 1181 (1989)].
[3] A. L. Vikharev, A. M. Gorbachev, A. V. Kim, and A. L. Kolysko, Fiz. Plazmy **18**, 1068 (1992) [Sov. J. Plasma Phys. **18**, 554 (1992)].
[4] V. G. Brovkin, Yu. F. Kolesnichenko, and D. V. Khmara, *Ball Lighting in the Laboratory* [in Russian], Khimiya, Moscow (1994).
[5] L. P. Grachev, I. I. Esakov, G. I. Mishin, and K. V. Khodataev, Zh. Tekh. Fiz. **64**(1), 74 (1994) [Tech. Phys. **39**, 40 (1994)].
[6] L. P. Grachev, I. I. Esakov, K. V. Khodataev, and V. V. Tsyplenkov, Fiz. Plazmy **18**, 411 (1992) [Sov. J. Plasma Phys. **18**, 216 (1992)].
[7] M. V. Kurlin, Ya. I. Panova, V. V. Pasynkov, and V. N. Tairov, *Radio-Frequency Electronic Materials* [in Russian], Sudostroenie, Leningrad (1969), 343 pp. .
[8] Yu. P. Raizer, *Gas Discharge Physics*, Springer-Verlag, Berlin–New York (1991).
[9] A. D. MacDonald, *Microwave Breakdown in Gases*, Wiley, New York (1966) [Russian trans., Mir, Moscow (1969), 206 pp.].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# The influence of electron emission from the collector on the potential distribution within a Knudsen diode with surface ionization in the underneutralized regime

V. I. Sitnov and A. Ya. Énder

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

(Submitted December 31, 1996)
Zh. Tekh. Fiz. **68**, 37–47 (April 1998)

This paper gives a theoretical treatment of the distribution of potentials within a Knudsen diode with surface ionization in the underneutralized regime in the presence of electron emission from the collector surface. A method is derived to calculate the potential distribution. It is shown that if the emission from the collector is sufficiently strong, spatially oscillatory (''wavelike'') potential distributions do not form; instead, a continuous transition occurs from a distribution with a virtual anode to one with a virtual cathode. Particular attention is focused on the neighborhood of the transition point from one of these distributions to the other. © *1998 American Institute of Physics.* [S1063-7842(98)00804-6]

Thermionic converters of thermal to electrical energy have their highest efficiency in the Knudsen regime, when the mean free path of electrons is larger than the interelectrode gap. In this regime, the emitter is simultaneously a source of electrons and ions (the electrons are generated by thermionic emission, while the ions arise from surface ionization), and the thermionic converter is a Knudsen diode with surface ionization. When there is no electron emission from the collector surface, the ideal current–voltage (I–V) characteristic of a Knudsen diode with surface ionization consists of two segments: a retarding segment, over which the current increases exponentially with increasing collector potential, and a saturation segment, over which the current is independent of the collector potential. At the boundary between these two segments, the external voltage $U$ equals the difference between the emitter work function $\varphi_E$ and the collector work function $\varphi_C$, i.e., $U=\varphi_E-\varphi_C$.

In Ref. 1, Dobretsov optimized a thermionic converter in the Knudsen regime without taking into account the special features of the potential distribution in the interelectrode gap. In Ref. 2, Babanin *et al.*, calculated the optimal parameters of a Knudsen diode with surface ionization using calculations of the potential distribution in which the collector current was neglected. As in Ref. 1, they took into account the collector current by adding it to the forward current obtained from a self-consistent calculation. They showed that a thermionic converter reaches its highest efficiency when the emission currents from the emitter and collector surfaces are comparable.

In a thermionic converter using Cs as a filler, the cesium vapor acts both to neutralize the electron space charge and to lower the electrode work functions. Because the optimal emitter work function is achieved only at rather high cesium vapor pressures, the Knudsen regime can be reached only for small interelectrode gaps ($d\approx 10$ $\mu$m). For wider gaps ($d\approx 1$ mm), the Knudsen regime can be reached in thermionic converters that use Cs–Ba as a filler. In these devices, barium vapor is used as an independent regulator of the emitter work function. It is worth noting that optimal values of the barium vapor pressure can only be reached when the barium heat source temperature is 1000 K or higher, and the collector temperature should be even higher than this. A hot collector coated with barium adsorbed on the surface possesses a rather high emissivity, comparable to that of the emitter, and the emission from the collector can have a substantial effect on the potential distribution. Thus, the investigation of self-consistent potential distributions that include the reverse current from the collector is an important problem.

In a Knudsen diode with surface ionization, ions flowing in from the emitter neutralize the electron space charge emitted by both the emitter and collector. The calculation of self-consistent solutions for the potential distribution in the presence of three groups of particles is complicated by the fact that these solution depend on a large number of parameters. This problem was first solved by McIntyre in Ref. 3; however, he did not attempt any systematic calculations. In this paper we propose a method for calculating and analyzing the potential distribution for a Knudsen diode with surface ionization in the underneutralized regime, assuming that the collector emission is unbounded. We briefly described some results of these calculations previously in Ref. 4.

**1.** One of the important parameters used to characterize a Knudsen diode with surface ionization is the degree of neutralization:

$$\gamma=\frac{n_i^+(0)}{n_e^+(0)},\tag{1}$$

where $n_i^+(0)$ and $n_e^+(0)$ are the densities of ions and electrons emitted from the emitter surface. Decreasing the emitter work function while keeping all the other parameters unchanged (for a Cs–Ba thermionic converter this can be done experimentally by increasing the barium vapor pressure) will decrease $\gamma$. As long as $\gamma>1$ (the overneutralized regime), the saturation current increases with decreasing $\gamma$.

The transition to the underneutralized regime ($\gamma<1$) is marked by the appearance of a retarding barrier for electrons near the emitter, which limits the growth in the electron current. A characteristic of this regime is the existence of a potential well for ions. Depending on how efficiently this well is populated by ions in the underneutralized regime, as $\gamma$ decreases the saturation current is either unchanged (complete occupation) or decreases. In Ref. 5 the authors advance the point of view that the potential well is fully occupied. However, in their paper they did not include boundary effects, which lead to expulsion of ions from the well. Moreover, nonuniformity of the surface of a real emitter can lead to the removal of ions, since spots with large work functions will be effective ion absorbers. This effect is especially important in the neighborhood of $\gamma=1$, where the well depth is comparable to the scatter in the values of the work functions of different spots.

In the calculations described in Refs. 6–9 it was assumed that the potential wells were not populated by ions. In this case it was found that a virtual cathode appears near the emitter, consisting of a barrier that retards electrons followed by an accelerating potential fall. Calculations of the structure of this virtual cathode showed that its height increases with decreasing $\gamma$, and consequently the directed electron velocity increases. In Refs. 10 and 11 a method was developed for diagnostics of the plasma in a Knudsen diode with surface ionization using a transverse magnetic field; in Ref. 12 it was shown experimentally that in the underneutralized regime there is in fact a high virtual cathode, and that the degree of occupation of the potential well with ions is close to zero. Therefore, in calculating the potential distribution in a Knudsen diode with surface ionization and an emitting collector one can neglect the trapping of ions in the potential well.

**2.** Let us first pause to discuss the basic methods for calculating potential distributions within a Knudsen diode with surface ionization without including electron emission from the collector, and the results that follow. Calculations of this kind have been described in a number of papers.[6,8,13–15] It is known that for sufficiently large values of the interelectrode gap, the potential distribution in a Knudsen diode with surface ionization is determined by two external parameters: the degree of neutralization and the dimensionless collector potential, and that monotonic potential distributions, spatially oscillating (‘‘wavelike’’) potential distributions, or distributions with a virtual cathode can occur in the gap. In the underneutralized regime, the minimum value of the potential for a nonmonotonic potential distribution is reached at a point close to the emitter, where the electric field is zero. The most detailed discussion of the structure of the virtual cathode and wavelike potential distributions, and their effect on the I–V characteristics, is given in Ref. 6. As an example, Fig. 1a shows several I–V characteristics for $\gamma=0.01$, while in Fig. 1b potential distributions are plotted for a number of points on these I–V characteristics. In Fig. 1a, the current is plotted in units of the emission current from the emitter:

$$j_e^+(0)=en_i^+(0)(2kT_E/\pi m)^{1/2}, \tag{2}$$

and in Fig. 1b the distance $z$ from the emitter is measured in units of the Debye radius $\lambda_D$ and the potential $\Phi$ relative to the emitter potential is plotted in units of $kT_E/e$. Here and in what follows, we will use the following dimensionless coordinate and potentials:

$$\zeta=z/\lambda_D, \quad \eta=e\Phi/kT_E, \quad u=eU/kT_E, \tag{3}$$

where $m$ is the electron mass, $k$ is Boltzmann's constant, $e$ is the electron charge, $T_E$ is the emitter temperature, and $U$ is the external voltage.

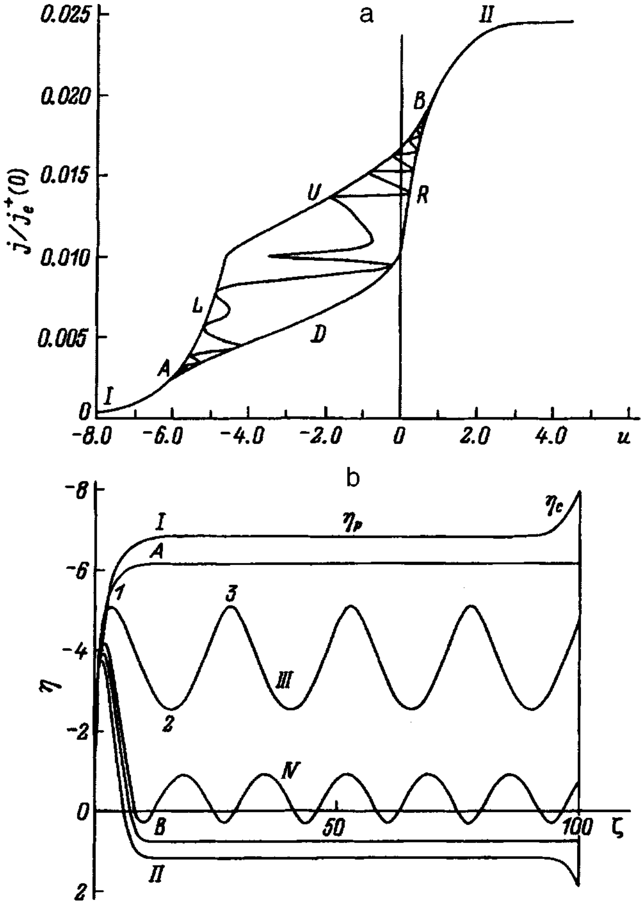

FIG. 1. Example of a current–voltage (I–V) characteristic (a) and potential distributions (b) at certain points on this I–V characteristic in the underneutralized regime of a Knudsen diode with surface ionization and a nonemitting collector. The value of the interelectrode gap is $d=100\lambda_D$.

The Debye radius is defined in terms of the emitter temperature and characteristic current density $j$ from the relation

$$\lambda_D=(kT_E/2\pi)^{3/4}(ej)^{-1/2}m^{-1/4}. \tag{4}$$

At a temperature $T_E=2000$ K and $j=1$ A/cm$^2$ the Debye radius $\lambda_D=4.6$ $\mu$m. In Fig. 1 the quantity $j_B$, i.e., the conduction current density corresponding to the boundary of a virtual cathode or a wavelike potential distribution (point $B$ in Fig. 1a), was used as the characteristic current for a given $\lambda_D$. The dimensionless potentials of the collector, plasma, and point of minimum potential are denoted by $\eta_C$, $\eta_p$, and $\eta_m$, respectively. In Fig. 1b, $\eta_m$ corresponds to the peak of the virtual cathode, or the minimum closest to the emitter for a wavelike potential distribution. In the underneutralized regime, the electron conduction current is determined by the value of $\eta_C$ only for large negative values of the collector potential, while in the rest of the potential range it is determined by the minimum value of the potential in the interelectrode gap. At large negative collector potentials the potential distribution is monotonic (curve *I* in Fig. 1b). Point *A* marks the transition to a wavelike potential distribution, while at point *B* this wavelike potential distribution becomse a distribution with a virtual cathode.

An important feature of the Knudsen regime is the fact that the potentials at characteristic points of the potential distribution can be found without completely solving the self-consistent system of Vlasov equations. Thus, when a virtual cathode is present two conditions hold that allow us to determine $\eta_p$ and $\eta_m$: quasineutrality in the neighborhood of the central plateau:

$$n_i|_{\eta=\eta_p}=n_e|_{\eta=\eta_p} \tag{5}$$

and average neutrality over the external part of the virtual cathode. This follows from Gauss's theorem and the vanishing of the field at the peak of the virtual cathode and in the plateau region:

$$\int_{\eta_m}^{\eta_p}(n_i(\eta)-n_e(\eta))d\eta=0. \tag{6}$$

In Eq. (6), the integration over coordinates has been replaced by an integration over $\eta$. In fact, from the Poisson equation $dE/dz=4\pi e(n_i-n_e)$ it follows that

$$\frac{1}{2}\frac{d(E^2)}{dz}=4\pi Ee(n_i-n_e).$$

By integrating this equation between the peak of the virtual cathode and the plasma plateau, and taking into account that $E=-d\Phi/dz$ and $E^2|_{\eta=\eta_m}=E^2|_{\eta=\eta_p}=0$, we obtain Eq. (6).

The functions $n_i$ and $n_e$ are proportional to $n_i^+(0)$ and $n_e^+(0)$, and depend not only on the local potential, but also on values of the potential at several characteristic points on the potential distribution to be found. When the velocity distribution of emitted particles is semi-Maxwellian, the functions $n_i$ and $n_e$ can be calculated analytically (the corresponding expressions for the more general case of a Knudsen diode with surface ionization and an emitting collector will be given below). In this case, the integral in Eq. (6) can be done analytically as well. This system of transcendental equations (5) and (6) can be used to find values of $\eta_m$ and $\eta_p$ for given $\gamma$ and $\eta_C$.

For a wavelike potential distribution, it is necessary to fix $\eta_m$ as a parameter, i.e., the value of the conduction current ($j=\exp(\eta_m)$). Then $\eta_2$ can be found from an equation analogous to Eq. (6) (Fig. 1b), and $\eta_3$ from the condition that the charge equals zero in the strip $(\eta_2,\eta_3)$. It is easy to find the boundaries of the region where wavelike potential distributions occur by setting $\eta_C=\eta_p$ for a monotonic potential difference or a virtual cathode. In Fig. 2 these boundaries are plotted as functions of $\gamma$: $\eta_A(\gamma)$ is the boundary between monotonic potential distributions and wavelike potential distributions, and $\eta_B(\gamma)$ is the boundary between wavelike potential distributions and a virtual-cathode distributions.

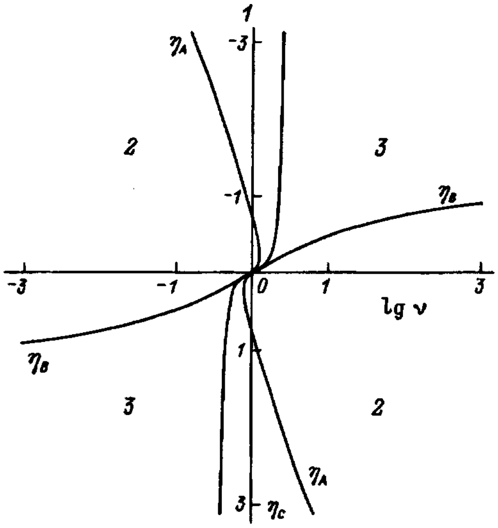

FIG. 2. Region of existence of various types of potential distributions in a Knudsen diode with surface ionization and a nonemitting collector: *1* — monotonic potential distribution, *2* — wavelike potential distribution, *3* — virtual cathode.

It can in fact be shown analytically that a cold collector cannot support a state with a quasineutral plasma when $\eta_C$ lies between $\eta_A$ and $\eta_B$. To show this, let us consider, e.g., a monotonic potential distribution with $\eta_C=\eta_p$, and then increase the value of $\eta_C$ somewhat. The potential plateau, if one exists, should be smaller than $\eta_C$. Then the condition of quasineutrality $n_i=n_e$ must hold at the boundary between the plasma and the cathode fall, and

$$\left.\frac{d(n_i-n_e)}{d\eta}\right|_{\eta=\eta_{p+0}}<0. \tag{7}$$

This relation follows from Gauss's theorem, taking into account the sign of the electric field in the cathode sheath region. The densities of ions and electrons for these specific assumptions are determined in the following way:

$$n_i=n_i^+(0)\mathrm{exers}(-\eta), \tag{8}$$

$$n_e=n_e^+(0)e^{\eta_p}\,\mathrm{exers}(\eta-\eta_p). \tag{9}$$

Here $\mathrm{exers}(x)=e^x(1-\mathrm{erf}\sqrt{x})$, and

$$\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt.$$

In taking the derivatives in Eq. (7) it is seen that a term $n_e^+(0)\exp(\eta_p)/(\pi(\eta-\eta_p))^{1/2}$ appears in $dn_e/d\eta$, which goes to $+\infty$ as $\eta\rightarrow\eta_{p+0}$, while the remaining terms are finite. Consequently, condition (7) is not satisfied, and solutions with a quasineutral plasma cannot exist for $\eta_p>\eta_A$. It can be shown analogously that no stationary regime can coexist with a quasineutral plasma for $d\eta_p<\eta_B$ as well. Consequently, in the interval $(A,B)$ only wavelike potential distributions can occur.

The authors of Refs. 6 and 7 made detailed calculations of wavelike potential distributions in a Knudsen diode with surface ionization and used them to construct the portion of the I–V characteristics in the range $\eta_A<\eta_C<\eta_B$. They showed that in this range there are many regions on the I–V characteristic with negative internal resistance. The I–V characteristic shown in Figure 1a, which corresponds to a wavelike distribution, is bounded by the curves *R, D, L, U.* Changing the magnitude of the interelectrode gap changes the I–V characteristics; specifically, increasing the gap increases the number of segments with negative resistance. However, for values of the gap larger than a certain minimum value ($\sim 10\lambda_D$) the I–V characteristics are confined within the region bounded by the curves *R, D, L, U.* Thus, the appearance of wavelike potential distributions for a certain set of values of the collector potential is a characteristic property of a Knudsen diode with surface ionization and a nonemitting collector.

In Ref. 3 McIntyre investigated the effect of electron emission from the collector theoretically for the first time, and showed that the solution to this problem depended on many parameters. In addition to the degree of neutralization $\gamma$ and the dimensionless collector potential $\eta_C$, he introduced two more dimensionless parameters: $\beta_C = n_e^-(\eta_C)/n_e^+(0)$ (where $n_e^-(\eta_C)$ is the density of electrons near the collector that leave its surface and move toward the emitter) and $\Theta^* = T_C/T_E$. It was noted that the ion and electron densities depend on both the local value of the potential $\eta$ and on the shape of the potential distribution. An analytical expression was obtained for the square of the electric field in the form of a certain function $g(\eta;\gamma,\eta_C,\beta_C,\Theta^*)$, where

$$dg(\eta)/d\eta = n_e^+(\eta) + \beta_C n_e^-(\Theta^*,\eta) - \gamma n_i(\eta).$$

The terms on the right-hand side of the equation are the densities of electrons emitted by the emitter and collector and the density of ions at a point with potential $\eta$. McIntyre developed two procedures to find the potential: in one, he directly integrated the Poisson equation, while in the other he integrated a first-order equation in which the derivative of the potential is written in terms of the function *g*. The results of these calculations are shown in Fig. 3. It is worth noting that these calculations were done for various collector potentials belonging to the same I–V characteristic, and for $\gamma$ close to the region of transition from overneutralized to underneutralized. Specifically, the external parameters were: $\gamma = 0.9$, $\beta_C = 0.5$, and $\Theta^* = 0.5$. A comparatively small value of the emission current from the collector was chosen [about $0.35 j_e^+(0)$].

McIntyre's calculations covered only a few examples and in fact serve only as rough estimates of the effect of emission from the collector. In discussing his results, he mentions[3] that the region of wavelike potential distributions turns out to be quite sensitive to changes in $\beta_C$. However, he obtained no systematic data on account of the large number of parameters and the complexity of the calculations. Such data are clearly necessary if the role of collector emission is to be understood, with particular attention paid to the case of strong emission.

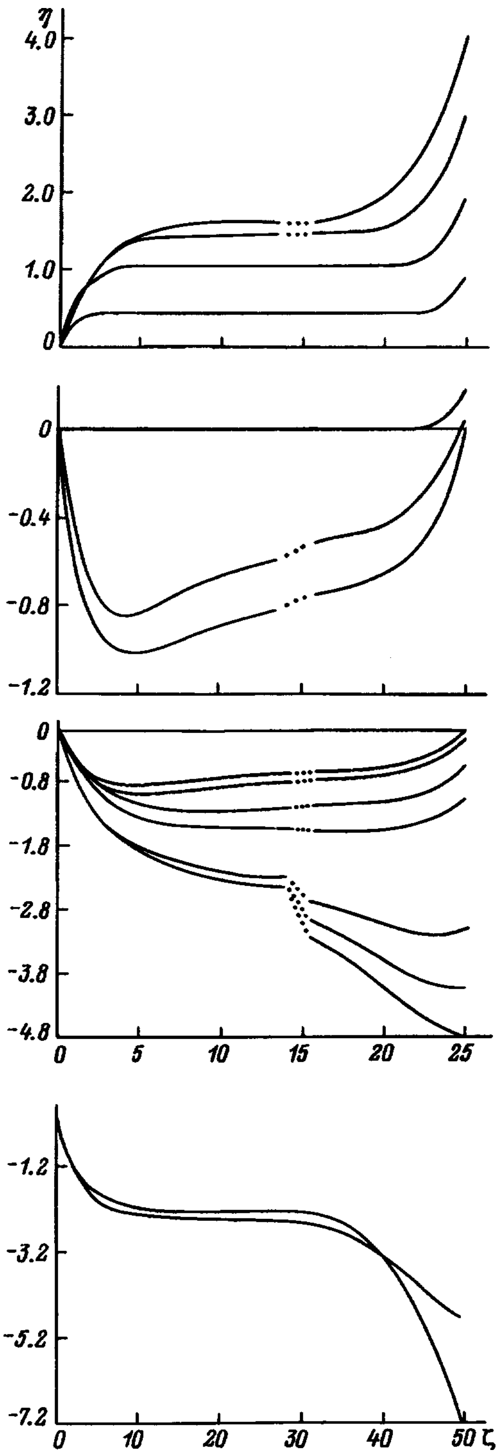

FIG. 3. Computed potential distributions as $\eta_C$ changes, based on Ref. 3.

**3.** Let us now consider a Knudsen diode with surface ionization in the underneutralized regime with an emitting collector. We will assume that emission from the collector is unbounded. This implies that emission from the collector is so large that the flux of electrons entering the plasma is limited only by the potential barrier near the collector, and not the collector emissivity. Discussion of this limiting case reveals how strongly emission from the collector can affect the potential distribution.

At rather large negative voltages, emission from the collector leads to the appearance of a negative current in the circuit — that is, a region of reverse currents appears on the I–V characteristics. An example of the potential diagram from this region is shown in Fig. 4. When there is no emis-

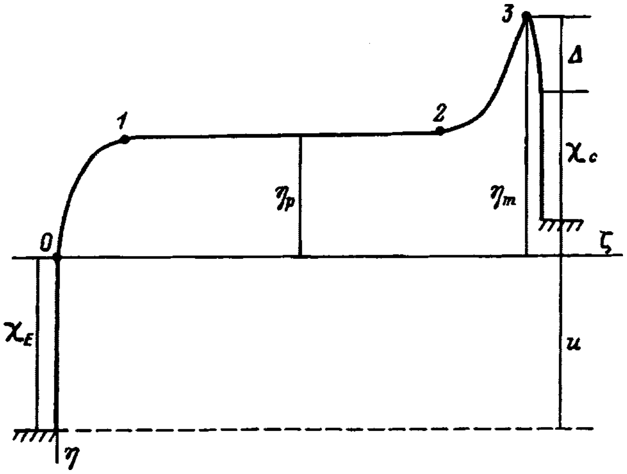

FIG. 4. Potential diagram for a Knudsen diode with surface ionization and a virtual anode: *1* — boundary of the cathode sheath, *2* — boundary of the virtual anode, *3* — top of the virtual anode; $\eta_1=\eta_2=\eta_p$, $\eta_3=\eta_m$.

sion from the collector, a large negative voltage will always correspond to a monotonic potential difference (Fig. 1b). Now, however, this is not so: near the collector, a virtual anode forms, and the minimum potential $\eta_m$ is now located near the anode at the virtual anode peak. In this potential diagram the following notation is used: $\chi_E=e\varphi_E/kT_E$ and $\chi_C=e\varphi_C/kT_E$, i.e., these quantities are dimensionless work functions of the emitter and collector respectively. It is clear from the figure that the effective work function of the collector $\chi_C^*$ equals

$$\chi_C^*=\chi_C+\Delta. \tag{10}$$

Here $\Delta$ is the height of the potential barrier for electrons emitted by the collector. The density of charged particles at an arbitrary point between the emitter and the virtual anode is determined as follows:

$$n_i(\eta)=n_i^+(0)\cdot F_i(\eta);\quad n_{eE}=n_e^+(0)\cdot F_{eE}(\eta,\eta_m);$$
$$n_{eC}=n_e^-(\eta_m)\cdot F_{eC}(\eta,\eta_m). \tag{11}$$

Here $n_e^-(\eta_m)$ is the density of electrons expelled by the collector and arriving at the virtual anode, evaluated at the peak of the latter (at the point $\eta_m$). As a unit for measuring the density of electrons emitted by the collector, we choose the quantity $n_e^-(\eta_m)$ rather than $n_e^-(\eta_C)$ (see Ref. 3), which is natural when the emission from the collector is unbounded. The functions $F$ are labeled so as to indicate the origin of the corresponding particles. Since all the potentials are measured in units of $kT_E/e$, the function $F_{eC}(\eta,\eta_m)$ depends on the ratio of the emitter and collector temperatures. For a fixed value of this ratio the quantity $n_e^-(\eta_m)$ is determined by the minimum potential in the interelectrode gap and does not depend on the collector work function. As $\chi_C$ varies, redistribution of the potential takes place only within a narrow layer between the top of the virtual anode and the collector.

Let us assume that the emitted particles have a semi-Maxwellian distribution with respect to velocity and temperature of the corresponding electrode. Then the functions $F_i$, $F_{eE}$, and $F_{eC}$ take the following form:[6]

$$F_i(\eta)=\mathrm{exers}(-\eta);$$
$$F_{eE}(\eta,\eta_m)=2e^{\eta}-e^{\eta_m}\mathrm{exers}(\eta-\eta_m);$$
$$F_{eC}(\eta,\eta_m,\Theta)=\mathrm{exers}((\eta-\eta_m)\Theta). \tag{12}$$

Here $\Theta=T_E/T_C$. The condition for quasineutrality of the plasma can be written as follows:

$$F(\eta_p,\eta_m)=0, \tag{13}$$

where

$$F(\eta_p,\eta_m)=\gamma F_i(\eta_p)-F_{eE}(\eta_p,\eta_m)$$
$$-\beta F_{eC}(\eta_p,\eta_m,\Theta). \tag{14}$$

Here the functions $F_i$, $F_{eE}$, and $F_{eC}$ are determined from Eq. (12) for $\eta=\eta_p$, $\gamma$ is given by Eq. (1), and

$$\beta=\frac{n_e^-(\eta_m)}{n_e^+(0)}. \tag{15}$$

It should be noted that the parameter $\beta$ depends on $\eta_m$, and consequently on the potential distribution to be found. Equation (13) relates the potentials at points $\eta_p$ and $\eta_m$ (points *3* and *2* in Fig. 4). A second equation that relates the potentials at these points can be derived from the condition that the total charge equal zero in a layer between these points, analogous to the derivation of Eq. (6). These contributions to the charge we denote by $n_i^+(0)G_i(\eta_p,\eta_m)$, $n_e^+(0)G_{eE}(\eta_p,\eta_m)$, and $n_e^-(\eta_m)G_{eC}(\eta_p,\eta_m,\Theta)$; the function $G_s(\eta_p,\eta_m)$ is calculated by integrating the corresponding function $F_s$ between the limits $\eta_m$ to $\eta_p$. For an arbitrary interval $(A,B)$ we have

$$G_s(A,B)=\int_A^B F_s(\eta)d\eta. \tag{16}$$

The condition that the total charge equal zero in the layer $(\eta_p,\eta_m)$ has the form

$$G(\eta_p,\eta_m)=0, \tag{17}$$

where

$$G(\eta_p,\eta_m)=\gamma G_i(\eta_p,\eta_m)-G_{eE}(\eta_p,\eta_m)$$
$$-\beta G_{eC}(\eta_p,\eta_m). \tag{18}$$

It is easy to show that

$$\int_A^B \mathrm{exers}(\eta)d\eta=Ir(B)-Ir(A), \tag{19}$$

where $Ir(\eta)=\mathrm{exers}(\eta)+2(\eta/\pi)^{1/2}$.

It is clear from Eq. (12) that calculating the function $G$ reduces to calculating integrals like Eq. (19) between the points $\eta_p$ and $\eta_m$. In this case

$$G_i=-Ir(-\eta)\Big|_{\eta_m}^{\eta_p};\qquad G_{eE}=(2e^{\eta}-e^{\eta_m}Ir(\eta-\eta_m)\Big|_{\eta_m}^{\eta_p};$$
$$G_{eC}=\frac{1}{\Theta}Ir(\eta)\Big|_0^{(\eta_p-\eta_m)\Theta}. \tag{20}$$

In order to find $\eta_p$ and $\eta_m$ it is necessary to solve the system of transcendental equations (13) and (17). In this system $\beta$ cannot be treated as an independent parameter. Let us show that it can be expressed in terms of $\eta_m$ and the parameters $\chi_E$, $\Theta$ and $u$. To do so, we use several additional relations. From Eq. (10) and the potential diagram (Fig. 4) it follows that

$$\chi_C^* = \chi_E + u - \eta_m . \tag{21}$$

For electrons that enter the interval between emitter and the point $\eta_m$ (points *0* and *3* in Fig. 4), we have the following absolute values of the currents:

$$j_e^+(0) = AT_E^2 \cdot e^{-\chi_E}, \quad j_e^-(\eta_m) = AT_c^2 \cdot e^{-\Theta \chi_c^*}. \tag{22}$$

From this

$$\frac{j_e^-(\eta_m)}{j_e^+(0)} = \Theta^{-2} \exp(\chi_E - \Theta \chi_C^*). \tag{23}$$

If we take into account that electrons at the top of the virtual anode moving toward the emitter have a semi-Maxwellian velocity distribution with a temperature $T_C$, and that $j_e = en_e \cdot v_T$ for the semi-Maxwellian distribution, where the thermal speed of the electrons $v_T \sim T^{1/2}$ [see Eq. (2)], then using Eq. (15) we obtain for the left-hand side of Eq. (23)

$$\frac{j_e^-(\eta_m)}{j_e^+(0)} = \frac{\beta}{\sqrt{\Theta}}. \tag{24}$$

Finally, taking Eq. (21) into account, we have

$$\beta = \Theta^{-3/2} \cdot \exp((\eta_m - u)\Theta - \chi_E(\Theta - 1)). \tag{25}$$

Note that $\beta$ depends on $u$ in addition to the external parameters $\Theta$ and $\chi_E$. Now the system (13), (17) can be solved and the quantities $\eta_p$, $\eta_m$ found, and the $\beta$ can be found from Eq. (25). Knowing $\eta_m$ and $\beta$, it is easy to compute the current in the external circuit (without including the ion current)

$$\frac{j}{j_e^+(0)} = \frac{j_e^+(\eta_m) - j^-(\eta_m)}{j_e^+(0)}.$$

Here $j_e^+(\eta_m)$ is the fraction of electron current from the emitter that overcomes the virtual anode and reaches the collector. It is obvious that $j_e^+(\eta_m) = j_e^+(0)\exp(\eta_m)$. Taking Eq. (24) into account, we have

$$\frac{j}{j_e^+(0)} = e^{\eta_m} - \frac{\beta}{\sqrt{\Theta}}. \tag{26}$$

Thus, for a given $u$ the current is determined, i.e., we have found a point on the I–V characteristic.

The external parameters that determine the form of the I–V characteristic are $\gamma$, $\chi_E$, and $\Theta$. If for fixed values of these parameters we increase the voltage $u$ beginning with large negative values, the height of the virtual anode $(\eta_p - \eta_m)$ will decrease and eventually will go to zero. The potential of the plasma at which this takes place we denote by $\eta_\alpha$; for the case of a nonemitting collector it corresponds to the transition from a monotonic potential distribution to a wavelike potential distribution (point *A* on Fig. 1).

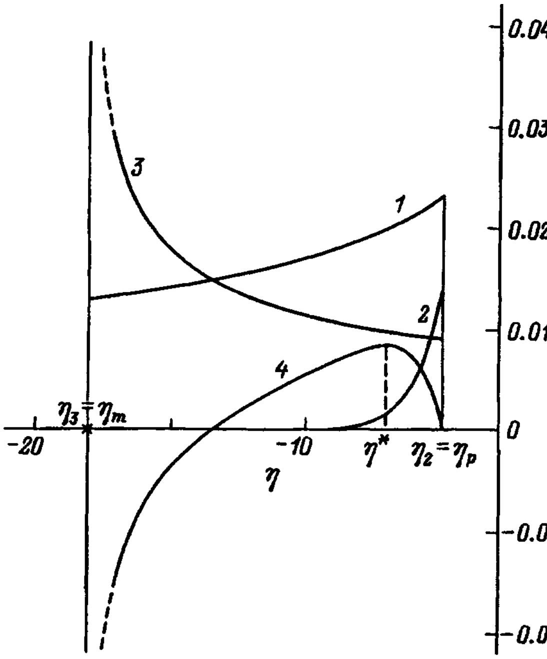

FIG. 5. $\eta$-dependence of the charged particle density over the external portion of the virtual anode $(\eta_2, \eta_3)$. $\chi_E = 15$, $\Theta = 2$, $u = -25$; *1* — $\gamma E_i$, *2* — $F_{eF}$, *3* — $\beta F_{eC}$, *4* — $(n_i - n_e)/m_e^+(0)$.

**4.** The parameters that characterize the point $\alpha$ can be found analytically. In the region between the points $\eta_2 = \eta_p$ and $\eta_3 = \eta_m$ (Fig. 4) the charge density, i.e., $n_i - n_e$, should change sign, otherwise, its integral over this layer cannot equal zero. Therefore, the function $n_i - n_e$ goes to zero at least twice: at the point $\eta_2$ (according to the condition of quasineutrality) and at some other point on the external portion of the virtual anode. As an illustration of this, in Fig. 5 the dependence of $n_i - n_e$ on $\eta$ is plotted on the interval $(\eta_2, \eta_3)$. Note that at the point $\eta_2$ we must have $d(n_i - n_e)/d\eta < 0$ in order for the quasineutral plasma potential to match the potential fall. Between the two points at which the function $(n_i - n_e)$ goes to zero there is a point where it reaches a maximum. Consequently, in the interval $(\eta_2, \eta_3)$ there necessarily exists a point (which we denote by $\eta^*$) at which the derivative $d(n_i - n_e)/d\eta$ vanishes. As the height of the virtual anode decreases we have $\eta_m \to \eta_p$, and all three points $\eta_3$, $\eta_2$, and $\eta^*$ approach infinitesimally close to one another as well as to $\eta_\alpha$.

Using the expressions for $F_s$ from Eq. (12), we obtain after differentiation

$$\left.\frac{d(n_i - n_e)}{d\eta}\right|_{\eta = \eta^*} = -\gamma F_i(\eta^*) - F_{eE}(\eta^*, \eta_m) - \Theta \beta F_{eC}(\eta^*, \eta_m, \Theta) + \gamma/\sqrt{-\pi\eta^*} - e^{\eta^*}/\sqrt{\pi(\eta^* - \eta_m)} + \beta/\sqrt{\Theta}\sqrt{\pi(\eta^* - \eta_m)}. \tag{27}$$

As $\eta_m \to \eta_p$, the last two terms in Eq. (27) become the main ones, each going to infinity as $\eta^* \to \eta_p$. In order for the derivative (27) to go to zero in this limit it is necessary that

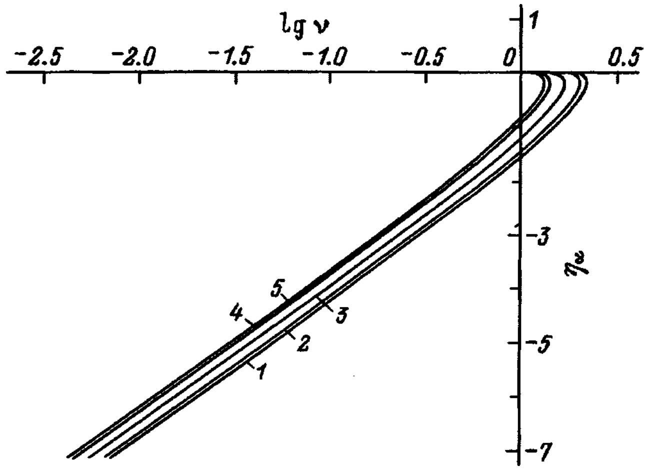

FIG. 6. The function $\eta_\alpha(\gamma)$ for various values of $\Theta$. $\Theta$: *1* — 2, *2* — 3, *3* — 10, *4* — 100, *5* — $\infty$.

TABLE I.

| $\gamma$ | $\Theta$ | $u_\alpha$ | $j_\alpha/j_e^+(0)$ | $\eta_\alpha$ | $\beta_\alpha$ |
|---|---|---|---|---|---|
| 0.2 | 1.6 | −7.23192 | 0.0133037 | −3.50178 | 0.0238305 |
| | 1.8 | −8.53668 | 0.0137910 | −3.47281 | 0.023128 |
| | 2.0 | −9.57041 | 0.0159099 | −3.44767 | 0.022500 |
| | 2.2 | −10.4809 | 0.0177439 | −3.42558 | 0.021932 |
| | 2.4 | −11.1016 | 0.0193525 | −3.40593 | 0.021415 |
| | 2.6 | −11.6834 | 0.0207790 | −3.38830 | 0.020941 |
| | 2.8 | −12.1785 | 0.0220556 | −3.37236 | 0.020503 |
| | 3.0 | −12.6048 | 0.0232071 | −3.35783 | 0.020098 |
| 0.1 | 1.6 | −7.52310 | 0.00519999 | −4.27827 | 0.010963 |
| | 1.8 | −8.88207 | 0.00634015 | −4.24990 | 0.010633 |
| | 2.0 | −9.95982 | 0.00731058 | −4.22529 | 0.010339 |
| | 2.2 | −10.83312 | 0.00814945 | −4.20366 | 0.010073 |
| | 2.4 | −11.55570 | 0.00888458 | −4.18444 | 0.0098314 |
| | 2.6 | −12.16270 | 0.00953578 | −4.16720 | 0.0096100 |
| | 2.8 | −12.67946 | 0.0101181 | −4.15160 | 0.0094060 |
| | 3.0 | −13.12447 | 0.0106492 | −4.13739 | 0.0092171 |
| 0.01 | 1.6 | −8.46136 | 0.000425978 | −6.78029 | 0.00089804 |
| | 1.8 | −9.99439 | 0.000519007 | −6.75286 | 0.00087023 |
| | 2.0 | −11.21116 | 0.000597756 | −6.72920 | 0.00084534 |
| | 2.2 | −12.19936 | 0.000665757 | −6.70840 | 0.00082294 |
| | 2.4 | −13.01724 | 0.000725295 | −6.68988 | 0.00080263 |
| | 2.6 | −13.70488 | 0.000778027 | −6.67324 | 0.00078408 |
| | 2.8 | −14.29088 | 0.000825034 | −6.65824 | 0.00076698 |
| | 3.0 | −14.79591 | 0.000867390 | −6.64455 | 0.00075119 |

$$\beta|_{\eta_m\to\eta_p}\to e^{\eta_\alpha}/\sqrt{\Theta}. \tag{28}$$

The condition of quasineutrality (14) for $\eta_m=\eta_p$ then simplifies to

$$\gamma F_i(\eta_\alpha)-e^{\eta_\alpha}-\beta=0. \tag{29}$$

Using Eq. (28) we obtain a relation between the potential at the point $\alpha$ and the parameters $\gamma$ and $\Theta$:

$$\gamma=\frac{e^{\eta_\alpha}}{F_i(\eta_\alpha)}(1+1/\sqrt{\Theta}). \tag{30}$$

From this it is easy to find the quantity $\eta_\alpha$ given $\gamma$ and $\Theta$. From Eqs. (25), (26), and (28) we have at the point $\alpha$

$$\frac{j_\alpha}{j_e^+(0)}=e^{\eta_\alpha}(1-1/\Theta),$$

$$u_\alpha=(\eta_\alpha-\chi_E)(1-1/\Theta)-\ln\Theta/\Theta,$$

$$\beta_\alpha=e^{\eta_\alpha}/\sqrt{\Theta}. \tag{31}$$

Note that $\eta_\alpha$, $j_\alpha$, and $\beta_\alpha$ depend only on $\gamma$ and $\Theta$, and that the only quantity that also depends on the emitter work function $\chi_E$ is the potential $u_\alpha$ of the external circuit at the point $\alpha$. The quantity $u_\alpha$ is conveniently written in the form of a sum of two terms $u_\alpha^{(1)}(\gamma,\Theta)+u_\alpha^{(2)}(\chi_E,\Theta)$. Here

$$u_\alpha^{(1)}(\gamma,\Theta)=\eta_\alpha(\gamma,\Theta)(1-1/\Theta)-\ln\Theta/\Theta,$$

$$U_\alpha^{(2)}(\chi_E,\Theta)=-\chi_E(1-1/\Theta).$$

In the limit $\Theta\to\infty$ Eq. (30) goes over to the function $\eta_A(\gamma)$ for a nonemitting collector (Fig. 2). Figure 6 shows curves $\eta_\alpha(\gamma)$ for various $\Theta$. Table I lists values of the remaining quantities at the point $\alpha$ for $\chi_E=15$, $\gamma=0.2$, 0.1, and 0.01, and a number of values of $\Theta$. In Fig. 7 the functions $u_\alpha^{(2)}(\chi_E)$ are plotted for several values of $\Theta$.

**5.** In order to investigate the region $u<u_\alpha$ it is necessary to solve the system of two transcendental Eqs. (13) and (17) using Eq. (25). Let us denote the left-hand side of Eq. (14) by $f_1(\eta_p,\eta_m)$ and that of Eq. (18) by $f_2(\eta_p,\eta_m)$. To solve this system numerically we used the generalized chord method for two dimensions. In the plane $\eta_p,\eta_m$ we choose three points that do not lie on a line, and at these points we calculate the quantities $f_1$. We then pass a plane through the corresponding three points in the space $f_1$, $\eta_p$, $\eta_m$, and find its line of intersection with the plane $f_1=0$. An analogous line is found for $f_2$ as well. Then the point of intersection of these lines is chosen as a new point on the plane $\eta_p$, $\eta_m$, and the procedure is repeated. This iterative process is continued until $f_1$ and $f_2$ become smaller than some given value (in these calculations, usually $10^{-6}$). This process converges quite rapidly, but in the neighborhood of $\eta_\alpha$ complications sometimes arise due to the closeness of $\eta_p$ and $\eta_m$.

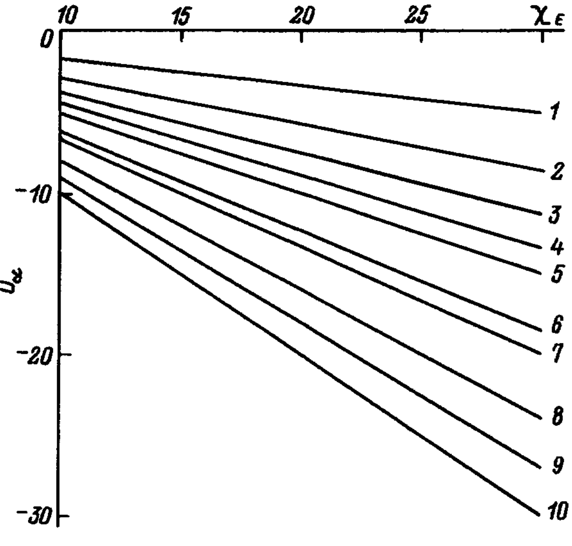

FIG. 7. The functions $u_\alpha^{(2)}(\chi_E)$ for various values of $\Theta$. $\Theta$: *1* — 1.2, *2* — 1.4, *3* — 1.6, *4* — 1.8, *5* — 2.0, *6* — 2.6, *7* — 3.0, *8* — 5.0, *9* — 10.0, *10* — $\infty$.

In order to study this region we can linearize the system of Eqs. (13), (17) in the neighborhood of a known solution $(\eta_\alpha, u_\alpha)$. Let us set $u = u_\alpha + \tilde{u}$. Using the notation $\tilde{\eta}_m = \eta_m(u) - \eta_\alpha$, $\tilde{\eta}_p = \eta_p(u) - \eta_\alpha$, $\tilde{\beta} = \beta(u) - \beta_\alpha$, and $\sigma = \eta_p - \eta_m = \tilde{\eta}_p - \tilde{\eta}_m$, we obtain after linearization

$$\tilde{\beta} = \beta_\alpha(\tilde{\eta}_m - \tilde{u})\Theta;$$

$$F_i = F_i(\eta_\alpha) + \left.\frac{\partial F_i}{\partial \eta}\right|_{\eta=\eta_\alpha} \tilde{\eta}_p;$$

$$F_{eE} = F_{eE}(\eta_\alpha) + \left.\frac{\partial F_{eE}}{\partial \eta}\right|_{\eta=\eta_\alpha} \tilde{\eta}_p + \frac{\partial F_{eE}}{\partial \sigma}\sigma;$$

$$F_{eC} = F_{eC}(\eta_\alpha) + \frac{\partial F_{eC}}{\partial \sigma}\sigma. \tag{32}$$

We then substitute Eq. (32) into Eq. (14) and group all the coefficients of $\sigma$:

$$\frac{\partial F_{eE}}{\partial \sigma} + \beta_\alpha \frac{\partial F_{eC}}{\partial \sigma} = \frac{e^{\eta_\alpha}}{\sqrt{\pi\sigma}} - \frac{\Theta\beta_\alpha}{\sqrt{\pi\Theta\sigma}} - \Theta\beta_\alpha = -e^{\eta_\alpha}\sqrt{\Theta}. \tag{33}$$

Here we have used the definition of $\beta_\alpha$ given in Eq. (31). Substituting Eq. (32) into the quasineutrality condition (13) and using Eq. (33) and the definitions of $\sigma$ and $\beta_\alpha$, we obtain

$$\gamma \left.\frac{\partial F_i}{\partial \eta}\right|_{\eta=\eta_\alpha} \tilde{\eta}_p - e^{\eta_\alpha}\tilde{\eta}_p - e^{\eta_\alpha}\sqrt{\Theta}(\tilde{\eta}_p - \tilde{u}) = 0.$$

Here $\partial F_i/\partial\eta|_{\eta=\eta_\alpha} = 1/\sqrt{-\pi\eta_\alpha} - F_i(\eta_\alpha)$. From this we find

$$\tilde{\eta}_p = k_p\tilde{u}; \quad K_p = \frac{\sqrt{\Theta}e^{\eta_\alpha}}{(1+\sqrt{\Theta})e^{-\eta_\alpha} - \gamma\partial F_i/\partial\eta|_{\eta=\eta_\alpha}}.$$

The coefficient $K_m$ as a function of $\tilde{\eta}_m(\tilde{u})$ in the neighborhood of the point $\alpha$ is determined from the condition that the charge within the exterior part of the virtual anode (17) equal zero. For this, the functions $F_s$ in integrals like Eq. (16) are expanded out to terms of first order of smallness, the integration is performed, and values of $F_s$ are substituted along with their derivatives at the point $\alpha$. It turns out that $K_m = K_p$.

We can continue the expansion of the functions $\eta_p$ and $\eta_m$ in powers of $\tilde{u}$. Let us consider the solution in the form

$$\eta_m = \eta_\alpha + K_m\tilde{u} + B_m\tilde{u}^2; \quad \eta_p = \eta_\alpha + K_p\tilde{u} + B_p\tilde{u}^2. \tag{34}$$

After some straightforward but tedious calculations we obtain

$$K_m = K_p = K = -\sqrt{\Theta}e^{\eta_\alpha}/A;$$

$$B_m = C/(2A) + \frac{2}{9}(B/A)^2;$$

$$B_p = C/(2A) + \frac{2}{3}(B/A)^2. \tag{35}$$

Here

$$A = \gamma\partial F_i/\partial\eta|_{\eta=\eta_\alpha} - (1+\sqrt{\Theta})e^{\eta_\alpha},$$

$$B = 2/\sqrt{\pi}(K(1-\Theta)+\Theta)e^{\eta_\alpha},$$

$$C = \Theta^{3/2}(K-1)^2 e^{\eta_\alpha} + K^2 e^{\eta_\alpha} - \frac{1}{2}\gamma K^2 \partial F_i^2/\partial\eta^2|_{\eta=\eta_\alpha}. \tag{36}$$

Thus, when the virtual anode disappears, the functions $\eta_m(u)$ and $\eta_p(u)$ have a common point and the same derivative at that point. As we move in the negative direction with respect to $u_\alpha$, these curves at first almost coincide, and then they diverge; the absolute value of $\eta_m$ increases somewhat more rapidly than $\eta_p$, i.e., a virtual anode of considerable height appears.

In order to investigate the evolution of the potential distribution as $u$ changes, we developed a special program. The primary task of this program is to calculate the quantities that are characteristic of the state $\alpha$. In the neighborhood of this point the values of $\eta_p$ and $\eta_m$ are found approximately using Eqs. (34)–(36). Then these values are refined using the procedure described above, which generalizes the chord method. Each time a new value of $u$ is chosen, approximate values of $\eta_p$ and $\eta_m$ are found by extrapolating the corresponding dependences at the previous points.

**6.** Let us now discuss how the potential distribution changes for $u > u_\alpha$. As we have said, when the collector is nonemitting the transition to this new potential arrangement eliminates distributions with a quasineutral plasma. The impossibility of matching the potential of the quasineutral plasma with that of the anode sheath stems from the conflicting requirements imposed on the behavior of charge in the sheath by the charged-particle equations of motion and by the the Poisson equation. This conflict strongly perturbs the potential distribution over the entire gap, leading to a transition to the wavy potential distribution (Fig. 1b, curves *III*, *IV*). This nonlocal effect of a change in the external voltage is characteristic of the Knudsen regime.

The presence of emission from the collector changes things considerably. As we already noted in deriving Eqs. (27), (28), in the neighborhood of the point $\alpha$ the reverse current from the collector is redistributed so as to compensate the infinite contribution to $\partial F_{eE}/\partial\eta$ as $\eta \to \eta_p$. As a result, the height of the virtual anode can go smoothly to zero. For $u > u_\alpha$ emission from the collector compensates the infinite *derivative* $F_{eE}/\partial\eta$. As a result, the conditions (7) for matching of the potential of the quasineutral plasma and the potential of the anode sheath can be satisfied, and it becomes possible for a virtual cathode to appear in the neighborhood of the point $\alpha$.

The potential diagram for this proposed potential distribution with a virtual cathode is shown in Fig. 8. Now the fundamental characteristic points of the distribution of potential turn out to be point *1* with potential $\eta_m$ and point *2* with potential $\eta_p$. The passage from virtual anode to virtual cathode leads to a change in the function $F_s$. On curve *1* the collector potential $\eta_C < 0$, and for $\eta_m < \eta < \eta_p$ we have[6]

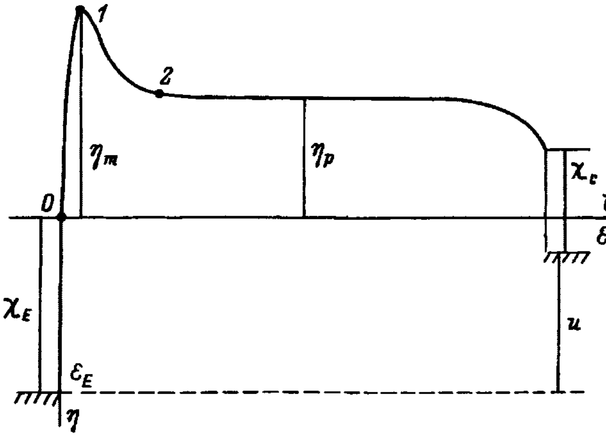

FIG. 8. Potential diagram for a Knudsen diode with surface ionization and a virtual cathode: *1* — top of the virtual cathode, *2* — boundary of the cathode sheath; $\eta_1=\eta_m$, $\eta_2=\eta_p$; $\varepsilon_E$ — Fermi level of the emitter, $E_C$ — Fermi level of the collector.

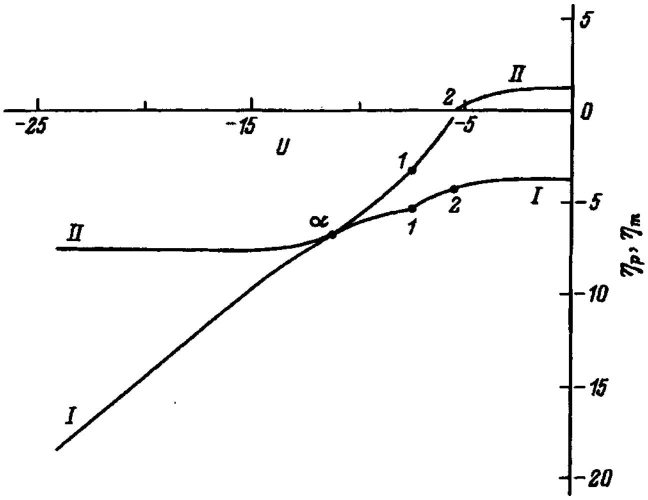

FIG. 9. The functions $\eta_m(u)$ (curve *I*) and $\eta_p(u)$ (curve *II*).

$$F_i(\eta)=\operatorname{exers}(-\eta);\qquad F_{eE}(\eta)=e^{\eta_m}\operatorname{exers}(\eta-\eta_m);$$

$$F_{eC}(\eta)=2e^{\Theta(\eta-\eta_m)}-\operatorname{exers}(\Theta(\eta-\eta_m)). \tag{37}$$

Note that $F_i(\eta)$ changes when $\eta_C>0$, and the ion density increases due to reflection from the anode sheath.

Let us continue this discussion for $\eta_C<0$. We now must satisfy Eqs. (13) and (17) using the functions $F_s$ of Eq. (37) and the functions $G_s$ calculated from them. For a virtual cathode, all the considerations invoked in deriving Eqs. (21)–(26) remain in force. In particular, Eq. (25) and the equation for the current (26), which relate $\beta$ to $\eta_m$ and the external parameters, are still valid. All the conclusions of Sec. 4 still hold, and in the limit as the height of the virtual cathode goes to zero a transition occurs to the same state $\alpha$ as the one in which there is no virtual anode.

Proceeding by analogy with our discussion of the virtual cathode region in Sec. 5, we investigated the curves $\eta_p(u)$ and $\eta_m(u)$ in the vicinity of the point $\alpha$. We showed that here $K_m=K_p=K$ as well. Furthermore, the values of the coefficients $K$, $B_m$, and $B_p$ do not change as we go from the virtual anode region to the virtual cathode region. Thus, at the point $\alpha$ the curves $\eta_m(u)$ and $\eta_p(u)$ are tangent to one another, and the point of tangency is approached very smoothly (Fig. 9). The quantities $\eta_p$ and $\eta_m$ were calculated in the neighborhood of the virtual cathode using the same scheme that we used in the neighborhood of the virtual anode. Figure 9 shows the functions $\eta_m(u)$ and $\eta_p(u)$ for $\gamma=0.01$, $\chi_E=15$, $\chi_C=7.5$, and $\Theta=2.0$. It is clear that for large negative $u$ the potential $\eta_m$ decreases almost linearly with decreasing $u$, while the plasma potential is practically unchanged. Because of the continuity of the derivative $d\eta_m/d\eta$ at the point $\alpha$, no kinks are observed in the I–V characteristics as we go from the virtual anode region to the virtual cathode region (Fig. 10). It is seen from the figure that a kink occurs at $u=\chi_E-\chi_C=7.5$, marking the point where reflection of ions from the anode potential fall begins.

**7.** Once the potentials at the characteristic points $\eta_p$ and $\eta_m$ for fixed external parameters and a given $u$ are found, it is easy to calculate the potential distribution in the interelectrode gap. For example, we can find the electric field at an arbitrary point $\eta$ on the segment between the top of the virtual anode (or virtual cathode) and the plasma plateau. We denote the dimensionless electric field at the point $\eta$ by $\varepsilon(\eta)$. If we choose the Debye radius Eq. (4) with characteristic current $j=j_e^+(0)$ as the unit of length, convert Eq. (7) to dimensionless variables, and integrate from $\eta_m$ to $\eta$, we obtain

$$-\frac{1}{2}\varepsilon^2(\eta)=\int_{\eta_m}^{\eta}(n_i(\eta)-n_e(\eta))d\eta$$

or

$$-\frac{1}{2}\varepsilon^2(\eta)=G(\eta,\eta_m)=\gamma G_i(\eta,\eta_m)-G_{eE}(\eta,\eta_m)$$

$$-\beta G_{eC}(\eta,\eta_m). \tag{38}$$

The function $G_s$ can be calculated analytically from Eq. (20) with $\eta_p$ replaced by $\eta$, and consequently the function $\varepsilon(\eta)$ can be found analytically. Using the known function $\varepsilon(\eta)$ and the relation $\varepsilon=-d\eta/d\zeta$ it is easy to find the potential distribution $\eta(\zeta)$ on the segment $(\eta_m,\eta_p)$ as well. Analogously we can construct the potential distribution on

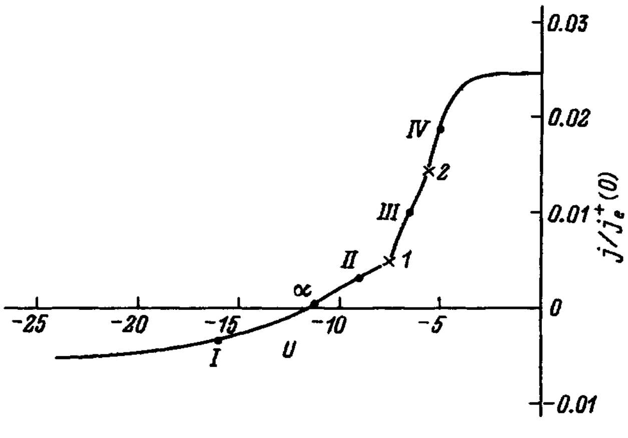

FIG. 10. Computed I–V characteristic of a Knudsen diode with surface ionization and an emitting collector.

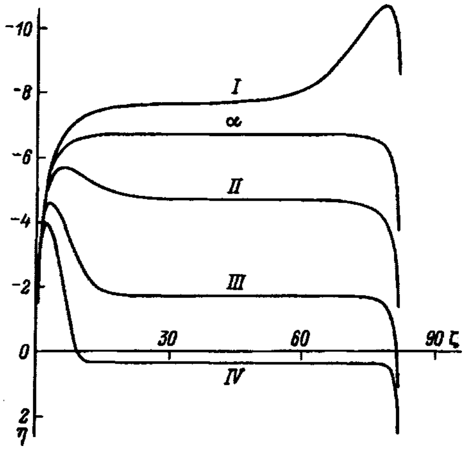

FIG. 11. Potential distribution at points labeled on the I–V characteristic of Fig. 10.

the other segments, changing only the expressions for the densities of charged particles. In the strongly underneutralized regime it is more convenient to choose the length unit as the Debye radius determined using the value of the saturation current $j_{\text{sat}}$ on the I–V characteristic as the characteristic current. Converting to this unit of length changes all the coordinates $\zeta$ by a factor of $(j_{\text{sat}}/j_e^+(0))^{1/2}$.

Figure 10 shows an example of an I–V characteristic computed for unbounded emission from the collector and the same external parameters used in Fig. 9, and Fig. 11 shows the potential distribution at several points of the I–V characteristic. Large negative external voltages $(u<u_\alpha)$ always lead to a potential distribution with a virtual anode (Fig. 11, curve *I*). As $u$ increases, the height of this virtual anode decreases, and at $u=u_\alpha$ it disappears (curve $\alpha$). For $u>u_\alpha$ a virtual cathode appears near the emitter (curve *II*), whose height increases with increasing $u$. As $\eta_C$ goes through zero (curve *III*), reflection of ions from the anode fall begins, and a kink appears on the I–V characteristic at the transition point (Fig. 10, curve *1*). Still another kink appears on the I–V characteristic at point *2* of Fig. 10, related to the transition of $\eta_p$ through zero (Fig. 11, curve *IV*) .

**8.** When there is no electron emission from the collector into the interelectrode gap, a quasineutral plasma cannot exist for potentials in the range $\eta_A<\eta_p<\eta_B$. The point $\alpha$ is analogous to point $A$ in a diode with an emitting collector. It is possible for a virtual cathode to coexist with a quasineutral plasma in the vicinity of $\eta_\alpha$ when $\eta_p>\eta_\alpha$ precisely because of emission from the collector. The corresponding term in the electron density $(n_e^-)$ ensures that $d(n_i-n_e)/d\eta$ can become negative as $\eta$ goes to $\eta_p$, making it possible to match the quasineutral plasma potential with that of the anode fall. Without a doubt, the disappearance of the region of wavelike potential distributions is the most interesting qualitative effect of emission from the collector.

A case of unbounded emission from the collector is encountered in thermionic converters with Cs–Ba fillers operated in the Knudsen regime. It was shown in Refs. 16 and 17 that such converters achieve their the highest power and efficiency at high emitter temperatures (∼2500 K). Under these conditions, the collector is heated to high temperatures, and possesses a high emissivity due to barium adsorbed on the surface. We can decide whether or not a collector possess ''unbounded'' emission by measuring the collector current as the emitter temperature changes. Increasing $T_E$ for fixed $\gamma$ and $T_C$ leads to an increase in the flux of ions and to a decrease in the height of the virtual anode. As a result, when the emissivity of the collector is high, increasing $T_E$ must increase the collector current. If, however, the current is limited by the emissivity of the collector, its value will not change. Experimental studies over a wide range of emitter temperatures show that the collector in a thermionic converter with Cs–Ba filler satisfies the criterion of unbounded emission. Thus, the assumption made in these calculations that the collector emission is unbounded turn out to be close to the real situation in thermionic converters with a Cs–Ba filler.

The authors are grateful to V. M. Babanin and V. I. Kuznetsov for helpful discussions, and to E. V. Yakovlev for help with certain calculations.

This work was carried out with the support of the Russian Fund for Fundamental Research (Project No. 97-02-18080).

[1] L. N. Dobretsov, Zh. Tekh. Fiz. **30**, 365 (1960) [Sov. Phys. Tech. Phys. **5**, 343 (1960)].

[2] V. I. Babanin, V. I. Kuznetsev, A. S. Mustafaev *et al.*, Zh. Tekh. Fiz. **48**, 754 (1978) [Sov. Phys. Tech. Phys. **23**, 444 (1978)].

[3] R. G. McIntyre, Proc. IEEE **51**, 760 (1963).

[4] A. Ya. Ender and V. I. Sitnov, in *Proceedings of the 23rd Intersociety Energy Conversion Engineering Conference* (Denver, 1988), Vol. 3, pp. 585–589.

[5] R. Ya. Kucherov and L. E. Rikenglaz, Zh. Tekh. Fiz. **28**, 1275 (1959) [Sov. Phys. Tech. Phys. **3**, 1185 (1959)].

[6] A. Ya. Énder, Candidate's Dissertation [in Russian] (Leningrad, 1972), 242 pp.

[7] V. I. Kuznetsov and A. Ya. Ender, *Third International Conference on Thermionic Electrical Power Generation*, Section F109 (Federal Republic of Germany, 1972).

[8] A. Ya. Ender, A. S. Mustafaev, V. I. Sitnov *et al.*, FTI Preprint Nos. 314, 315 [in Russian] (A. F. Ioffe Physicotechnical Institute, Academy of Sciences of the SSSR, Leningrad, 1971).

[9] V. I. Babanin, V. I. Kuznetsev, A. S. Mustafaev, and A. Ya. Énder, Zh. Tekh. Fiz. **42**, 2144 (1972) [Sov. Phys. Tech. Phys. **17**, 1695 (1972)].

[10] A. Ya. Énder, Zh. Tekh. Fiz. **38**, 1925 (1968) [Sov. Phys. Tech. Phys. **13**, 1546 (1968)].

[11] A. Ya. Énder, Zh. Tekh. Fiz. **40**, 551 (1970) [Sov. Phys. Tech. Phys. **15**, 711 (1970)].

[12] V. I. Babanin, M. B. Barabash, Yu. G. Kravinskiĭ *et al.*, Zh. Tekh. Fiz. **40**, 561 (1970) [Sov. Phys. Tech. Phys. **15**, 434 (1970)].

[13] R. G. McIntyre, J. Appl. Phys. **33**, 2485 (1962).

[14] W. Ott, Z. Naturforsch. A **22**, 1057 (1967).

[15] S. Kuhn, Plasma Phys. **23**, 881 (1981).

[16] A. Ya. Ender, V. I. Kuznetsov, V. I. Sitnov *et al.*, *Eleventh Symposium on Space Nuclear Power and Propulsion* (Albuquerque (USA), 1994), Vol. 2, pp. 861–867.

[17] V. I. Babanin, I. N. Kol'shkin, V. I. Kuznetsov *et al.*, *Second Satellite Conference on Nuclear Energy in Space: Physics of Thermionic Energy Converters* [in Russian] (Sukhumi, USSR, 1991), pp. 87–89.

Translated by Frank J. Crowne

# Minimum-wear characteristics for boundary friction between solids

Yu. P. Kozyrev and B. M. Ginzburg

*Institute for Problems in Mechanical Engineering, Russian Academy of Sciences, 119178 St. Petersburg, Russia*

(Submitted April 25, 1996; resubmitted April 15, 1997)
Zh. Tekh. Fiz. **68**, 48–52 (April 1998)

A new parameter is proposed for characterizing boundary friction between solids—the probability $Q$ that some contact spot will be converted to a wear particle as the real contact area shifts by one mean contact-spot diameter. A method is proposed for the phenomenological detection of optimum regimes of boundary friction, which are characterized by the minimum wear intensity of the sample material, and for measuring the corresponding quantity $Q=Q^{\mathrm{opt}}$. For babbitt, one of the most frictionless materials, the value of $Q^{\mathrm{opt}}$ is found to be $\approx 2\times 10^{-10}$. When data on the linear wear intensity $I_h$, the contact pressure $p_n$, and the hardness $H$ are available, $Q$ can be calculated for the given test conditions. Deviations of $Q$ from the value $Q^{\mathrm{opt}}$ (for a given material) can serve as a criterion for how closely a prepared surface structure approaches the optimal. © *1998 American Institute of Physics.* [S1063-7842(98)00904-0]

It is customary to sort the various friction mechanisms into three basic categories: dry, boundary, and liquid. In dry friction, the interaction of the bodies in contact is so strong that it often leads to scoring, which limits the use of dry friction in practical applications. Dry friction is also characterized by intense wear. In liquid friction, the bodies in contact are separated by a layer of liquid lubricant and are not in contact with one another, which leads to almost no wear. Only in boundary friction, which is intermediate between liquid and dry friction, can we conveniently observe and study the run-in processes that are characteristic for various materials in contact.[1–3] These processes give rise to changes in the surface structure of the contacting bodies at the macro- and microscopic levels that lead to minimum wear intensity for a given set of friction conditions.

In determining the wear intensity of materials, the test conditions and materials that make up the bodies in contact are usually chosen in such a way that the wear on one of the bodies can be neglected, while the second body whose wear is being measured is used as a test sample. This leads to the following question: does there exist some material characteristic that will indicate an optimum regime of boundary friction (and probably an optimum structure of a sample surface) and accordingly the minimum wear intensity of the sample material under various test conditions (pressure, type of lubricant, etc.)? A second question also arises: can we characterize wear for a set of materials with the help of this quantity and rank these materials based on its corresponding values?

In this paper we will attempt to answer these questions by evaluating those properties of materials that are relevant to the problem of optimizing surface structures so as to minimize wear intensity during boundary sliding friction.

The magnitude of the linear wear intensity $I_h$ is often cited as a characteristic of wear resistance.[4] However, this quantity depends on the test conditions, and tends to measure the wear resistance of machine components and mechanisms made of a given material rather than the wear resistance of the material itself. A quantity that is somewhat more useful in this regard is the so-called wear factor[5] $\Phi$, defined by the relation $\Phi=I_h/p_n$, where $p_n$ is the nominal contact pressure. However, this quantity is also unsatisfactory, since it can be completely different under different external conditions and at different stages of the wear process. Moreover, it does not take into account the differing wear resistances of materials with different strengths or hardnesses. Samples made of harder materials may have better wear resistance than softer materials, but this tells us nothing about their performance under optimal conditions, for which their wear resistance could be even higher (the wear resistance is inversely proportional the wear intensity or the wear factor).

In Ref. 6, Archard established that the following relation holds for dry sliding friction:

$$I_h=kp_n, \tag{1}$$

where $k$ is a constant coefficient.

In the literature it has been reported more than once that a certain range of contact pressures exists for which the quantity $I_h$ is a linear function of $p_n$, not only for dry friction but also for other kinds of friction. Constancy of $k$ within this pressure range implies that the wear mechanism is not changing. On the other hand, the pioneering studies by Khrushchov[7] of sliding friction between nonconforming bodies in lubricated contact revealed that over a wider range of $p_n$ friction with lubrication is characterized by a more complicated function $I_h=f(p_n)$ (Fig. 1), with intense wear observed on segment *I* ($p_b<p_n<p_a$), followed by a sharp falloff in the wear on segment *II* ($p_0<p_n<p_b$) and practically no wear at all on segment *III* ($p_n<p_0$).

Unfortunately, the literature contains very few systematic studies of boundary friction between two bodies of the same kind in contact over a wide range of external conditions. In Ref. 8, Markov showed, in agreement with our own processing of Khrushchov's data[7] (see below), that wear is

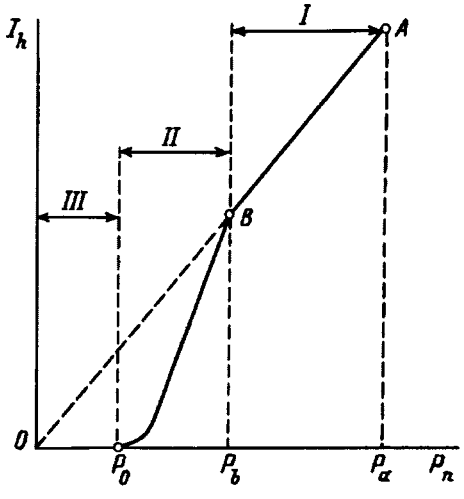

FIG. 1. A plot of linear wear intensity versus pressure under conditions of lubrication according to Ref. 7.

observed on segment *III* (although it is small). To illustrate this, let us use the results of Ref. 7 mentioned above, which are unusual in that their experimental conditions involve the same two materials. In these experiments, a rotating cylindrical ingot of St-45 steel was placed in contact along a generator of the cylinder with a stationary flat sample made of babbitt B83 (83% Sn, 11% Sb, 6% Cu). The lack of mating of the contact surfaces led to a constantly changing contact area, allowing us to obtain values of $I_h$ for babbitt over a wide range of pressures.

Using the tabulated data given in Ref. 7, we were able to construct the function $I_h(p_n)$. For two such functions $I_h(p_n)$ (Fig. 2, curves *1, 2*) we observed a bend at the lowest investigated values of $p_n$, corresponding to a transition to a wear regime in which the curve is directed toward the coordinate origin. Note that the functions $I_h(p_n)$ shown in Fig. 2 are plotted in logarithmic coordinates in order to include all the experimental data. In ordinary coordinates, straight-line segments like *III* that point to the coordinate origin become more extended. We will assume that it is these segments of type *III* that characterize the approach of the surface structure to optimal under the conditions of our tests. This assumption is confirmed by the fact that these type-*III* segments lie on the same straight line for both experimental curves (line *4* in Fig. 2) and are in fact continuations of each other.

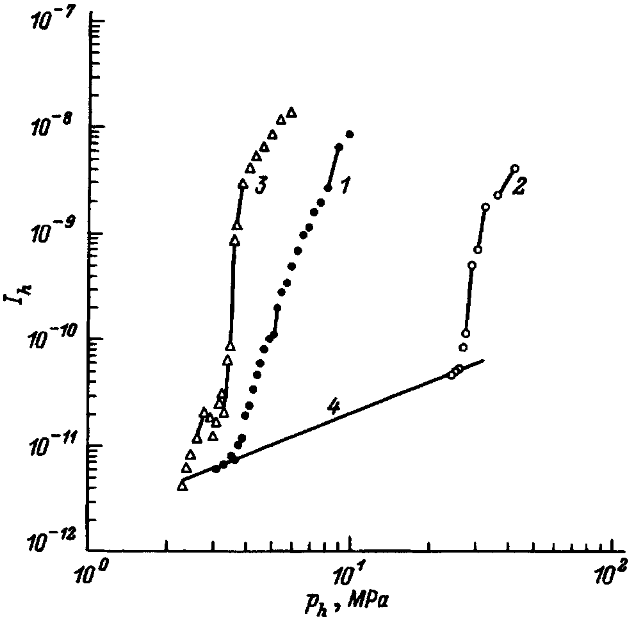

FIG. 2. Plots of linear wear intensity versus pressure constructed from data presented in Ref. 7. Material — babbitt B83, rotating ingot — St-45 steel. The rate of sliding was 3 m/s. Loads in *N*: *1,3* — 100, *2* — 300. Lubricants: *1,2*—Avtol, *3*—kerosene; *4*—the function $I_h = kp_n$.

Processing of the results of another series of tests of the same materials gives curve *3* in Fig. 2. This curve exhibits no type-*III* segment, probably because we did not obtain any lower values of the pressure (and in this case the test lasted 12 days!). However the lowest point on curve *3*, corresponding to the smallest value of $I_h$ obtained, falls right on the same straight line *4* (Fig. 2). Thus, the straight line *4* in Fig. 2 describes a function of type (1):

$$I_h^{\text{opt}} = K^{\text{opt}} p_n . \tag{2}$$

The same coefficient of proportionality $K^{\text{opt}}$ characterizes the minimum observed wear intensities for all three experimental curves shown in Fig. 2 obtained under various initial loads and various types of lubrication. As we mentioned above, constancy of this proportionality coefficient in situations described by functions like (1) implies that the wear mechanism does not change. Consequently, we may assume that when we fix all the friction conditions except the dependence on load and type of lubricant, the wear mechanism is the same for the wear curves shown in Fig. 2 (at least for curves *1* and *2*) at the minimum values of the external pressure. Since in this case the wear is minimal, we will assume that the corresponding surface structure of babbitt under friction is optimal for the conditions of boundary friction. This optimal structure does not depend either on the applied load or the type of lubricant. Its wear is described by the function (2) and is characterized by a linear wear intensity for optimal boundary friction $I_h^{\text{opt}}$.

Let us attempt to understand the physical meaning of the coefficient $K^{\text{opt}}$ in expression (2). For further mathematical calculations we will follow Ref. 6 and assume that the wear particle volume $V_d$ is proportional to the area of the contact spot. Then we have

$$V_d \approx \Delta h d^2 , \tag{3}$$

where $d$ is the diameter of the contact spot and $\Delta h$ is the height of the wear particle.

Let us denote by $\alpha$ the number of contact spots converted into wear particles and detached as the surface is worn down when the real contact area shifts by the average diameter of a contact spot. Then the probability that any given contact spot is converted into a wear particle (in principle, several wear particles can be formed out of one contact spot) equals

$$Q = \alpha / n , \tag{4}$$

where $n = A_r/d^2$ is the overall number of contact spots that make up the real contact area $A_r$.

Combining expressions (3) and (4), we find the specific volume wear during a shift by one contact spot diameter:

$$V_e = Qn\Delta h d^2. \qquad (5)$$

It is well known that the magnitude of the specific wear intensity is determined by relation[4]

$$i_h = V_e/A_r d. \qquad (6)$$

Substituting Eq. (5) into Eq. (6), we obtain

$$i_h = Q\frac{\Delta h}{d}. \qquad (7)$$

On the other hand, in steady-state wear regimes the magnitude of the intensity of linear wear $I_h$ is connected with the specific wear intensity $i_h$ by the following relation[4]

$$I_h = i_h\frac{A_r}{A_n}, \qquad (8)$$

where $A_n$ is the nominal contact area.

If, as in Ref. 9, we assume that

$$A_r \approx \frac{3F_n}{H}, \qquad (9)$$

where $F_n$ is the normal load and $H$ is the Brinell hardness of the material, then taking Eqs. (6) and (8) into account, we obtain

$$I_h \approx \frac{3\Delta h Q}{dH} p_n. \qquad (10)$$

In Eq. (10) we have the relation $\Delta h/d \leqslant 1$; equality to unity corresponds to the case of a wear particle with in the form of a lump, which is observed quite often in the wear of metals. During the wear of polymer materials the wear particles can have platelet shapes, and then the ratio $\Delta h/d$ is considerably less than 1.

Thus, comparing Eqs. (2) and (10) allows us to express the coefficient $K^{\mathrm{opt}}$ in terms of a number of physical quantities with well-defined meanings, i.e., the macroscopic characteristics of the wear process $I_h$, the microcharacteristics of the wear particles $\Delta h$ and $d$, and the coefficient $Q$ that characterizes the stochastic character of the process.

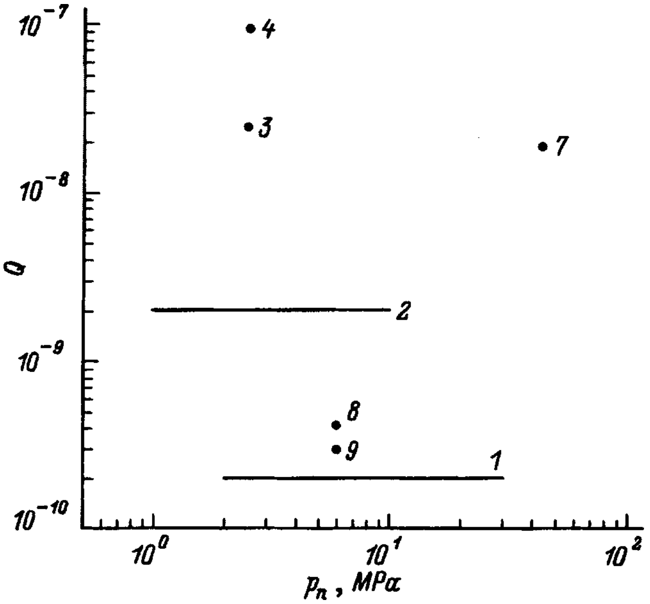

FIG. 3. Dependences of the probabilities of formation of a wear particle, $Q$, on the pressure $p_n$ for wear-resistant antifriction materials. The curves and experimental points are labeled with the number of the material in Table I.

It is natural to assume that for the optimum regime of boundary friction the probability of forming a wear particle from a given contact spot is a minimum. Let us denote it by $Q^{\mathrm{opt}}$. In order to determine the value of $Q^{\mathrm{opt}}$ for babbitts we used the plots given in Fig. 2. Using curve *4* (Fig. 2), we can compute the coefficient $K^{\mathrm{opt}}$, and use it in Eq. (10). Then if we set $\Delta h = d$ and use well-known data on the hardness of babbitt, we can compute the quantity $Q^{\mathrm{opt}} = K^{\mathrm{opt}}H/3$. This value, which appears in Fig. 3 as a horizontal line parallel to the abscissa, can serve as a single reference parameter to characterize the minimum wear for well tested antifriction materials such as babbitts.

Let us compare our value of $Q$ with values computed from literature data on other materials. To do so, we use Eq. (10) in those cases where values of $I_h$, $p_n$, and the hardness are given in the literature for tested materials with high wear

TABLE I. Characteristic values of the probability $Q$ for wear-resistant antifriction materials.

| No. of curve | Friction pair | Brinell Hardness, MPa | Pressure, MPa | Lubricant | Probability $Q$, $10^{-10}$ | Ref. |
|---|---|---|---|---|---|---|
| 1 | Babbitt B-83 — St-45 steel | 300 | 2–30 | Avtol-10, kerosene | 2 | 7 |
| 2 | FGM (Teflon-4 + graphite + MoS$_2$) — 3X13 steel | 60 | 1–10 | Without lubric. | 20 | 10 |
| 3 | BrOTsS5-5-5 bronze — St-45 steel | 600 | 2.5 | Industrial oil | 250 | 4 |
| 4 | St-45 steel — BrOTsS5-5-5 bronze | 2300 | 2.5 | ″ ″ | 950 | 4 |
| 5 | BrOF-10-1 bronze — 20X steel | 900 | – | I-30 | 99 | 11 |
| 6 | BrOTsS5-5-5 bronze — ShKh-15 steel | 600 | – | I-30 | 100 | 11 |
| 7 | Titanium carbide coating on St-20 steel — St-45 steel | $3\times10^4$* | 43 | I-40A | 185 | 12 |
| 8 | Pig iron — tin-coated ring | 3000 | 6 | Motor oil | 3 | 4 |
| 9 | Pig iron — chrome-coated ring | 3000 | 6 | ″ ″ | 4.2 | 4 |

*Value of the microhardness. The values of $Q$ are determined for the first-named material in each of the friction pairs listed in the first column.

resistance. Data for such materials used in machine components are listed in Table I, and the results of calculations plotted on the same Fig. 3. In all cases the computed values of $Q$ are at least no smaller than the value of $Q^{\text{opt}}$ for babbitts, and for the cast-iron bushings of motors with high wear resistance they are close to $Q^{\text{opt}}$ for babbitts.

This last fact suggests that the minimum wear is characterized by roughly the same values of $Q^{\text{opt}}$ for other materials as well.

We can regard the value of $Q^{\text{opt}}$ for babbitts as a reference characteristic relative to which we can rank other materials. Conversely, if we determine values of $Q^{\text{opt}}$ for other materials, then if the values of $Q$ calculated under given test conditions always deviate from the value $Q^{\text{opt}}$ for a given material, the deviation can serve as an estimate of the extent to which the prepared surface structure departs from optimal.

Thus, in this paper we have proposed to use the value of the probability of forming a wear particle $Q^{\text{opt}}$ under conditions of boundary friction in an optimum steady-state regime as a reference parameter to characterize the minimum wear of materials.

It is assumed that this quantity depends only slightly on the nature of the bodies in contact when these bodies come in contact under conditions of sliding friction, and that it has a value of the order of $\sim 10^{-10}$.

## CONCLUSIONS

1. Analysis of the literature shows that by measuring the linear wear intensity $I_h$ over a wide range of variation of the contact pressure $p_n$ (for example, by using nonconforming bodies in contact), and by varying the initial loads and types of lubricants, we can reach regimes of friction characterized by minimum wear under given experimental conditions. Relations of type $I_h^{\text{opt}} = K^{\text{opt}} p_n$ are valid ways to characterize these regimes, regardless of the dependence on initial loading or type of lubricant. We have used the quantity $K^{\text{opt}}$ to compute the probability $Q^{\text{opt}}$ of forming a wear particle from a contact spot in this regime (when the real contact area shifts by the average diameter of the contact spot), and we propose to use this quantity as a reference parameter to characterize the minimum wear of a given material and the optimum structure of its surface.

2. The quantity $Q^{\text{opt}}$ for babbitt, one of the most frictionless materials, was found to be $2 \times 10^{-10}$. We assert that this value depends only slightly on the nature of the bodies in contact for the case of sliding friction, and it has a value of the order of $\sim 10^{-10}$.

3. In any frictional regime for which data on the linear wear intensity $I_h$, the contact pressure $p_n$, and the hardness $H$ are available, we can compute the probabiltiy $Q$ for forming a wear particle from a contact spot. The deviation of values of $Q$ calculated under these experimental conditions from the value of $Q^{\text{opt}}$ for this material can serve to estimate how much the prepared surface structure departs from the optimum.

4. At this time the value of $Q^{\text{opt}} = 2 \times 10^{-10}$ obtained for babbitt is the smallest value of $Q$ computed for any material according to published data.

This work was carried out within the framework of the Russian Science and Technology Program ''Fullerenes and Atomic Clusters'' (Project ''Tribol'').

[1] B. E. Klamecki, Wear **58**, 325 (1980).
[2] B. I. Kostetskiĭ, Dokl. Akad. Nauk Ukr. SSR Ser. A, No.4, pp. 52–57 (1984).
[3] B. I. Kostetskiĭ, Trenie Iznos **14**(4), 73 (1993).
[4] I. V. Kragel'skiĭ, M. M. Dobychin, and V. S. Kombalov, *Principles of Computations for Friction and Wear* [in Russian], Mashinostroeniye, Moscow (1977), 440 pp.
[5] V. S. Kombalov, *Estimating the Tribotechnical Properties of Surfaces in Contact* [in Russian], Nauka, Moscow (1983), 324 pp.
[6] J. E. Archard, J. Appl. Phys. **24**, 981 (1953).
[7] M. M. Khrushchov, *Studies of the Performance of Alloys Used in Ball Bearings and Trunnions* [in Russian], USSR Acad. Sci. Publ., Moscow, Leningrad (1946), 252 pp.
[8] D. P. Markov, Trenie Iznos **14**(3), 601 (1993).
[9] F. P. Bowden and D. Tabor, *The Friction and Lubrication of Solids* (Clarendon Press, Oxford, 1954), 450 pp.
[10] V. S. Kombalov, Trenie Iznos **2**(1), 134 (1981).
[11] V. V. Alisin, in *Interdisciplinary Problems in Science and Engineering* [in Russian], Vol. 15 (GOSNITI, Moscow, 1980).
[12] V. P. Bulatov, B. M. Ginzburg, Yu. P. Kozyrev *et al.*, Trenie Iznos **15**(6), 1009 (1994).

Translated by Frank J. Crowne

# Nonuniform distribution of absorbed energy in high-resistance materials excited by an electron beam

V. F. Shtan'ko, V. G. Glybin, and V. M. Tolmachev

*Tomsk Polytechnical University, 634004 Tomsk, Russia*

(Submitted November 28, 1996)
Zh. Tekh. Fiz. **68**, 53–59 (April 1998)

Experimental results are presented on the changes in the optical characteristics of lithium fluoride induced by an electron beam with time-varying density and pulse energies close to the threshold for destruction of the material. The spatial distribution of color centers is investigated, especially near breakdown channels. Mechanisms for nonuniform accumulation of defects are discussed, along with the fundamental causes of the inhomogeneous energy distributions induced by the high-current electron beam. Concrete results of calculations of the field intensity distribution in LiF crystals during irradiation are presented, based on models of ''uniform'' and nonuniform charging of the sample. An abrupt increase in the electric field intensity is predicted near the breakdown channel.
[S1063-7842(98)01004-6]

## INTRODUCTION

Subjecting materials to pulsed high-current electron beams can produce levels of excitation high enough to cause damage,[1,2] making it possible to obtain concentrations of short-lived products sufficient for the detection and investigation of the mechanisms of defect formation in solids.[3,4] In addition to generating electrons and holes, high-current electron beams also create strong electric fields in the irradiated sample, and dynamic mechanical stresses that can influence both the efficiency of creation and decay of primary radiolysis products and post-irradiation processes. Because the degree to which these factors affect a material has been inadequately studied, it is difficult to include them correctly in analyzing experimental results.

When alkali halide crystals are irradiated by high-current electron beams with energy densities above their damage thresholds (0.1 J/cm$^2$; see Refs. 5 and 6) the electric field of the injected charge initiates multichannel electrical breakdown. This in turn leads to deviations in the spatial distribution of radiation defects and luminescence,[7] gives rise to fine structure on the acoustic pulse, and is one of the mechanisms for brittle fracture in insulators.[8] How effectively the electric-discharge channel dissipates energy depends on the density of the high-current electron beam, the properties of the material, and the irradiation geometry.

In this paper we analyze theoretically the phenomena that occur in the high-resistance material LiF excited by a high-current electron beam, either in the absence of or in the course of multichannel electric breakdown. Our primary focus is on the space–time distribution of the space charge, the profile of energy release, and the electric field intensity in this material. We also investigate experimentally the effect of breakdown on the spatial distribution and structure of defects in LiF.

## SPATIAL DISTRIBUTION OF ABSORBED ELECTRONS

In calculating the distribution of absorbed electrons in a crystal, we make use of one of several ''coarsened'' collision schemes[9] based on the Moller–Bethe angular distribution. The energy loss due to ionization is determined from the Bethe–Bloch formula, while the average ionization potential for the constituent elements is based on the Sternheimer formula.[10] The shape of the accelerating voltage pulse and current density $J(t)$ for a GIN-400 based accelerator were given in Ref. 11. The current density $J_{\max}$ at the maximum of the excitation pulse varied from 1 to 100 A/cm$^2$. At the pulse maximum, the accelerating voltage was 260 kV. In these calculations, a variable-density excitation pulse of duration 23 ns was divided up into $n$ monoenergetic pulses (in our case $n=23$ with a duration of $\Delta t=1$ ns or $n=230$ for $\Delta t=0.1$ ns) with electron energy $E_i$ and current density $J_i=G_iJ_{\max}$ in the $i$th pulse, where $G_i$ is the current pulse shape normalized to have a maximum of unity.

In the range of electron energies of interest to us (40–260 keV) the distribution of absorbed electrons with respect to distance into the crystal is independent of the initial energy of the incident electrons within error limits ($<5\%$). When plotted in reduced coordinates, this distribution follows a universal curve $F(U_{j,i})$ normalized to have a unit maximum ($F_{\max}=1$). Here $U_{j,i}=Z_j/R(E_i)$ is the reduced $j$th plot coordinate for the $i$th pulse, and $R(E_i)$ is the total range of an electron with energy $E_i$ calculated by integrating the Bethe–Bloch function.[9] In Ref. 11, we published detailed tables of $F(U_{i,j})$, $R(E_i)$, and the fraction of reflected electrons $P_r=N_{\text{ref}}/N_{\text{in}}$, which is a weak function of $E_i$ in the energy range 40–300 keV, for ten materials (KBr, KJ, KCl, KF, NaCl, NaI, LiF, CaF$_2$, CdS, and polymethylmethacrylate [PMMA]). Then the distribution of the volume charge density $\rho(Z_i,t_m)$ at the time of the excitation pulse has the following form in absolute coordinates for various densities and arbitrary pulse shape ($m<n$):

$$\rho(Z_i,t_m)=\left[J_{\max}(1-P_r)\Delta t\sum_{j=1}^{m}G_iF(U_{i,j})/R(E_i)\right]\Bigg/\left[\int_0^1F(U)dU\right]. \qquad (1)$$

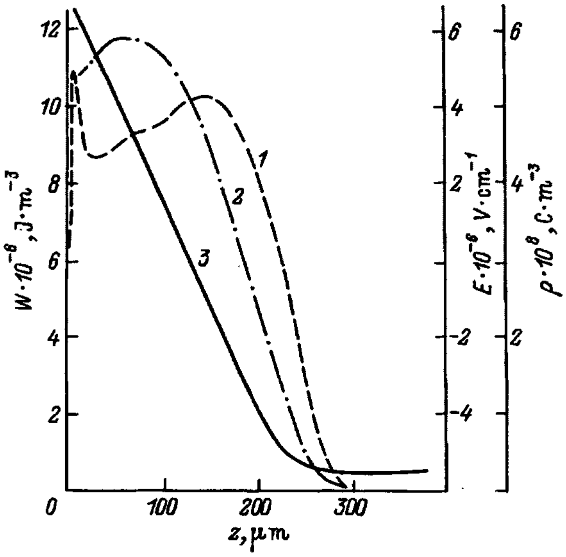

FIG. 1. Distribution of injected charge density $\rho(z)$ (*1*), absorbed energy density $W(z)$ (*2*), and electric field $E(z)$ (*3*) after irradiation by a high-current electron beam.

Equation (1) allows us to solve the self-consistent problem of including the retarding effect of the space charge in the crystal on the incident high-current electron beam by replacing $R(E_i)$ in Eq. (1) by $R(E_i-\Phi_{i-1})$, where $\Phi_{i-1}$ is the potential at the surface of the sample after $i-1$ single pulses. Including the retarding effect of the field leads to a shift in the maximum of the volume charge density toward the irradiated sample surface with increasing beam current density, and the beam electrons can be completely reflected from the sample at the end of the excitation pulse. In all the calculations that follow, $\rho(Z,t)$ incorporates the retarding effect of the field of the injected charge.

## CALCULATION OF THE ENERGY-RELEASE PROFILE FIELD IN A UNIFORM SAMPLE EXCITED BY A HIGH-CURRENT ELECTRON BEAM

For a uniform planar sample with known $\rho(z)$, in order to find the potential distribution $\Phi(z)$ and field intensity $E(z)$ in quasi-steady-state it is sufficient to solve the tridiagonal matrix Poisson equation by the method of double cyclic reduction[12] in the form

$$\Phi_{j-1}-2\Phi_j+\Phi_{j+1}=\Delta^2 W_j^0=W_j \tag{2}$$

with boundary condition $\Phi_0=\Phi_j=0$, where $0<j<J$; $\Delta$ is the mesh step (in our case, $\Delta\leqslant 10$ μm); $W_j^0=\rho_j/(\varepsilon_0\varepsilon)$; $J=L/\Delta$; $L$ is the distance between surfaces with zero potential; and $\varepsilon$ and $\varepsilon_0$ are the dielectric constant of the sample and the permittivity of free space, respectively.[1)]

Calculations show (Fig. 1) that for $L=R_{\max}$ (where $R_{\max}$ is the maximum range of the electrons) the maximum field intensity is reached at the surface being irradiated.

An important feature of the interaction between the high-current electron beam and the sample is the complicated space–time variation of absorbed energy in the latter. Figure 2 shows the time dependence of the volume density of energy released in LiF at various distances $z$ from the irradiated surface (the integrated energy release profile at the end of the pulse of irradiating electrons is shown in Fig. 1).

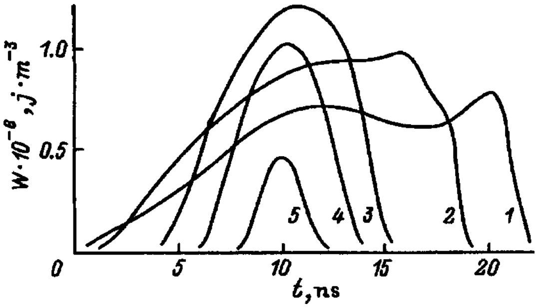

FIG. 2. Time dependence of energy density absorbed in LiF. $z$, μm: *1* — 10, *2* — 50, *3* — 150, *4* — 200, *5* — 250.

These results show that the irradiation time of a layer of crystal decreases as its distance from the irradiated surface increases. The volume energy density, and accordingly the rate of generation of nonequilibrium carriers, is maximum for a layer located at a depth of 150–200 μm. Thus the stopping of a high-current beam leads to space–time nonuniformity of the energy release and at the same time to the onset of an electric field.

The distribution $E(z)$ depends on the sample thickness. As this thickness increases, the field intensity at the irradiated surface $E(0)$ increases, reaching a limiting value of $E_\sigma=Q_{\text{inj}}(1-P_r)/(\varepsilon\varepsilon_0)$, where $Q_{\text{inj}}$ is the surface density of injected charge, while at the unirradiated surface the field $E(L)$ decreases. We have proposed empirical expressions to determine $E(0)$, $E(L)$ and $V_e=dE_{\text{sef}}/dt$ (with an error of $<10\%$) at any time during the excitation pulse when $L>R_{\max}$ in a sample with both surfaces grounded, or when the distance between the surface of the sample and a surface with $\Phi=0$ is much smaller than the sample thickness (a plus corresponds to $E(0)$, a minus to $E(L)$, and $Z_{\text{eff}}$ is the effective charge of atomic nuclei in the crystal),

$$E_{\text{sef}}=(1-P_r)Q_{\text{inj}}[(L-2R_{\max}/\text{Ln}(2Z_{\text{eff}}))/L\pm 1]/(2\varepsilon\varepsilon_0), \tag{3}$$

$$V_e(t)=(1-P_r)[(L-2R_{\max}/\text{Ln}(2Z_{\text{eff}}))/L\pm 1]J(t)/(2\varepsilon\varepsilon_0), \tag{4}$$

The numerical values of $E_{\text{sef}}$ and $V_e$ computed from Eqs. (3) and (4) for $J_{\max}=20$ A/cm$^2$ ($Q_{\text{inj}}=2.4\times 10^{-1}$ C/cm$^2$) are more than an order of magnitude lower than the criterion for streamer breakdown.[13] In contrast to this prediction, breakdown is observed experimentally[6,8] when $J_{\max}<20$ A/cm$^2$. The disagreement arises from the lack of criteria for initiation of breakdown in this model, the fact that the change in the charge distribution due to the nonuniformity of the electron emission from the surface was not included, etc.

## FIELD CALCULATIONS UNDER CONDITIONS OF NONUNIFORM CHARGING OF THE SAMPLE

According to Ref. 14, the emission current from one channel is large enough to completely neutralize the charge injected into the volume from an irradiated surface region with a radius of up to 150 μm. Therefore, our calculations were done for a cylindrical sample with radius $R_k=150$ μm, in which a single breakdown channel develops perpendicular to the surface and the degree of neutralization $G_k$ can reach 1.

For a system with radial symmetry, the Poisson equation in cylindrical coordinates has the form

$$d^2\Phi/dz^2 + d^2\Phi/dr^2 + (1/r)d\Phi/dr = -\rho(r,z)/(\varepsilon\varepsilon_0),$$

$$d^2\Phi/d\varphi^2 = 0; \quad 0 < r < R_k;$$

$$0 < z < Z_k; \quad d\Phi/dr|_{r=0} = 0. \tag{5}$$

Approximating the operators $d^2/dz^2$, $d^2/dr^2$, $d/dr$ by difference operators on a mesh $I,J$ with steps $\Delta_j = Z_k/J$ and $\Delta_i = R_k/I$ (we will treat the general case where $\Delta_i \neq \Delta_j$ and $\Delta_j = H\cdot\Delta_i$, where $H$ is a scale factor), and assuming that the function $\rho_{i,j}$ is determined at each mesh point $(i,j)$, we find that Eq. (5) can be transformed into a five-point matrix equation

$$\Phi_{i,j+1} - 2\Phi_{i,j} + \Phi_{i,j-1} + H^2[(1+1/2i)\Phi_{i+1,j} - 2\Phi_{i,j} + (1-i)\Phi_{i-1,j}] = -\Delta_j^2\rho_{i,j}/(\varepsilon\varepsilon_0), \tag{6}$$

with boundary conditions $\Phi_{i,0} = \Phi_{i,j} = 0$; $\Phi_{i-1,j} = \Phi_{i+1,j}$.

For the boundary conditions chosen, the potentials $\Phi_{i,j}$ and $\rho_{i,j}$ can be expanded in Fourier sine series along the columns of the mesh $(0 < j < J)$:

$$\Phi_{i,j} = \sum_{k=1}^{J-1} \hat{\Phi}_i(k)\sin(\pi kj/J);$$

$$\rho_{i,j} = \sum_{k=1}^{J-1} \hat{\rho}_i(k)\sin(\pi kj/J), \tag{7}$$

where $\hat{\Phi}_i(k)$ and $\hat{\rho}_i(k)$ are coefficients of the Fourier series.

Substituting Eq. (7) into the five-point equation (6), we obtain

$$\sum_{k=1}^{J} \{H^2[(1+1/2i)\hat{\Phi}_{i+1}(k) - 2\hat{\Phi}_i(k) + (1-1/2i) \times \hat{\Phi}_{i-1}(k)] + [2\cos(\pi kj/J) - 2]\hat{\Phi}_i(k) + (\Delta_j^2/(\varepsilon\varepsilon_0))\hat{\rho}_i(k)\}\sin(\pi kj/J) = 0. \tag{8}$$

Because Eq. (8) holds at every point $j$, we can individually set the amplitude of each Fourier harmonic equal to zero. Then for each $k$ we have a tridiagonal matrix equation

$$A_i\hat{\Phi}_{i+1}(k) + B(k)\hat{\Phi}_i(k) + C_i\hat{\Phi}_{i-1}(k) = \hat{W}_i(k), \tag{9}$$

where $A_i = H^2(1+1/2i)$, $C_i = H^2(1-1/2i)$, $B(k) = 2(\cos(\pi kj/J) - 1 - H^2)$, $\hat{W}(k) = -[\Delta_j^2/(\varepsilon\varepsilon_0)]\hat{\rho}_i(k)$, and $\hat{\rho}_i(k) = (2/J)\Sigma_{j=1}^{J}\rho_i\sin(\pi kj/J)$. This equation is solved for $\{\hat{\Phi}_i(k)\}$ by the method of double cyclic reduction.[12] In addition, the potential and electric field are determined according to Eq. (7) at every mesh point $(i,j)$ at any time within the excitation pulse.

Figure 3 shows the calculated distributions of the space charge $\rho(z)$ and electric field $E(z)$ in the direction of incidence of electrons in the pulsed irradiation of a LiF crystal, plotted at a distance $r = 1$ $\mu$m from the breakdown channel when $G_k = 0.5$. The radius of the breakdown channel is $r' = 1$ $\mu$m, and the radius of the neutralized volume $R_k = 150$ $\mu$m. As the degree of neutralization increases to 1, the magnitude of $E(z)$ at the leading edge of the breakdown channel already exceeds the dynamic electrical breakdown strength of LiF, which is 3.2 MV/cm, during the first nanoseconds of irradiation.[15,16] An increase in $E(z)$ is also observed as the radius of the breakdown channel decreases. Thus, for $r' = 0.5$ $\mu$m and $G_k = 1$, the field intensity at the leading edge of the breakdown increases to 180 MV/cm, i.e., much higher than the electrical breakdown strength of LiF and approaching the value of the intrinsic crystalline field.

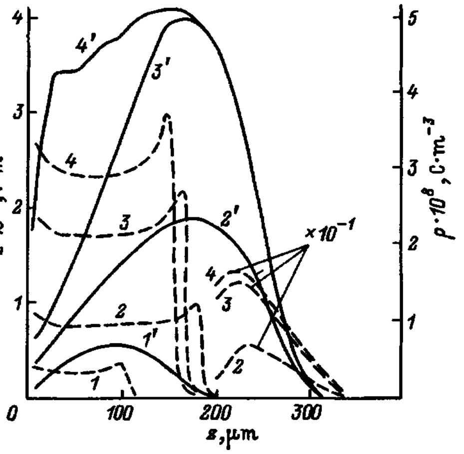

FIG. 3. Distribution of injected charge density (*1'–4'*) and electric field (*1–4*) at various times during a high-current electron beam pulse (in ns). *1,1'* — 5; *2,2'* — 10; *3,3'* — 15; *4,4'* — 23.

With distance from the breakdown channel in the radial direction the field intensity $E(r)$ first decreases rapidly near the channel and then more slowly at distances $r > (3-4)r'$. When $G_k = 1$ a high value of $E(r) > 10^6$ V/cm is observed for $r > 100$ $\mu$m from the breakdown channel.

These calculations indicate that for large values of $G_k$ and small breakdown channel radii $r'$ conditions for the initiation of electrical breakdown are satisfied even in the first nanoseconds of irradiation when $J_{\max} = 100$ A/cm$^2$, in agreement with experimental results.

## SPATIAL DISTRIBUTION OF DEFECTS IN LIF

The results given above imply that when an ionic crystal is irradiated by a high-current electron beam, the processes of thermalization of nonequilibrium electrons and holes, production and decay of excitons, and secondary conversion of primary structural defects all take place under the influence of a number of factors: the rate of generation of electrons and holes is nonuniform in time on account of the differing time dependences of the specific volume density of the energy released at different depths (Fig. 2), there is an electric field produced by the space charge, and there are mechanical stresses induced by pulsed heating of the region where the electron beam is stopped. In addition, irradiation by a high-current electron beam with an above-threshold current density leads to multichannel electric breakdown, which generates local dynamic stresses, a high electric field with space–

time nonuniformities (Fig. 3), a thermal field, and electromagnetic radiation emitted by the plasma in the breakdown channel.

In order to estimate the effect of these factors on the high-current electron beam, we studied the post-radiation spatial distribution of electronic color centers in LiF crystals. The choice of LiF as a system to study was dictated by its high mechanical strength, which allows multichannel breakdown to create visually observable tracks without brittle fracture, thereby ensuring that the coordinates of the region being probed can be determined. Moreover, the optical characteristics of the electron color centers created in LiF by irradiation are quite well-studied.[17] Whereas irradiation of LiF crystals at 295 K by a pulsed electron beam leads primarily to the generation of $F$ centers[18], more complicated $F$-aggregate centers are created by post-irradiation processes. For example, according to the mechanism proposed in Ref. 19, the formation of an $F_2$ center takes place in two stages: first an $F_2^+$ center forms when an $F$ center interacts with an anion vacancy, and then the new center captures an electron. This mechanism, whose existence in LiF was verified in Ref. 20, was also found to generate more complicated $F$-aggregate defects.[20] When crystals with a prior population of electron color centers are irradiated, the nature of the original centers can change as they interact with nonequilibrium electrons and holes.[21,22] In Ref. 17, Nahum proposed a mechanism in which $F_3^+$ and $F_3$ centers are generated via migration of a $F_2^+$ center and its interaction with $F$ centers. Hence, when LiF is irradiated at 295 K by a series of high-current electron beam pulses, several mechanisms participate in the generation of $F$-aggregate centers, the most effective of which are vacancy-related.

The optical absorption spectra and optical density $D(\lambda,x,z)$ were measured as a function of the coordinates $x,z$ of the region being explored using light at fixed wavelength $\lambda$ incident either normal to the irradiated surface, yielding $D(\lambda,x)$, or parallel to it, yielding $D(\lambda,z)$. Prior to these measurements, the LiF samples were irradiated by a train of electron pulses until multichannel breakdown tracks appeared, which were observed using a D11U11 microscope, and the samples were kept at 295 K until the color centers stabilized. The time evolution of a multichannel breakdown track in LiF was described in detail in Ref. 23.

The distribution $D(\lambda,z)$ was measured for two sample regions, the first containing a channel breakdown track (i.e., the probe beam intersects the channel) and the second without one (Fig. 4). Figure 5 shows portions of the optical absorption spectrum in the energy range 2.0–3.3 eV for crystal layers located at various distances $z$ from the irradiated surface. In layers far from the surface the spectrum is dominated by an absorption band with a maximum of 2.8 eV, which roughly coincides with the $F_2$ band. However, at $z=160$ $\mu$m this 2.8 eV band is broadened (its half-width is 0.24 eV as opposed to 0.176 eV for the $F_2$ band) due to overlap with $F_3^+$ center bands at 2.93 and 2.65 eV. The absorption spectrum is observed to become quite complex as the layer under study approaches the irradiated surface.

In order to analyze the spectral content of the 2.8 eV band, we used the Alentsev–Fok method.[24] We found that as $z$ decreases, the optical density in the bands at 3.3 eV (the $F_3$ band) and 2.26 eV (the $F_4$ band) increases. Moreover, we identified bands in the absorption spectra at 2.43, 2.5, and 3.02 eV for $z<70$ $\mu$m. According to Ref. 25, the band at 2.43 eV is due to $F_5$ centers. Based on our analysis of the spectral content of the absorption, we obtained spatial distributions of the real optical density with respect to $z$ for bands associated with $F_2$, $F_3^+$, and $F_4$ centers (Fig. 4). From this analysis we concluded that the closer the maximum of $D(\lambda,z)$ is to the irradiated surface, the more complicated the structure of the center is. When the probe beam intersected a breakdown track, we observed some decrease in the overall absorption over the spectral range of our measurements. In the immediate vicinity of a channel breakdown track we observed a relative increase in the absorption for optical bands caused by complex $F_3$, $F_3^+$, $F_4$, and $F_5$ centers, which are close to one another in energy.

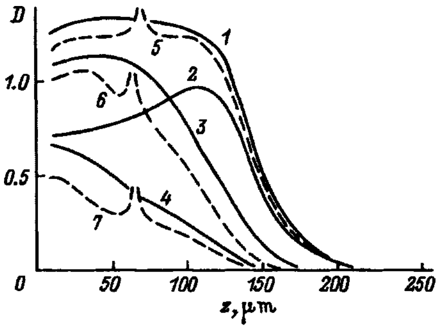

FIG. 4. Distribution of optical density in LiF in the direction along the normal to the irradiated surface. Probe beams do not (*1–4*) or do (*5–7*) intersect a breakdown track: *1,5* — for $\lambda=445$ nm; *2* — $F_2$; *3* — $F_3^+$; *2,7* — $F_4$ bands; *6* — for $\lambda=480$ nm.

The changes in the optical density at 2.8 eV revealed by measurements in the $x$ direction were considerably more complicated when the probe beam intersected diverging channel breakdown tracks (Fig. 6) located in a plane parallel to the irradiated surface. In this case, the value of the optical density exhibits well-defined modulations with alternating maxima between 35 and 40 $\mu$m. Furthermore, a comparison of the absorption spectra of regions near breakdown channels and at distances 50, 100, and 150 $\mu$m from them shows that

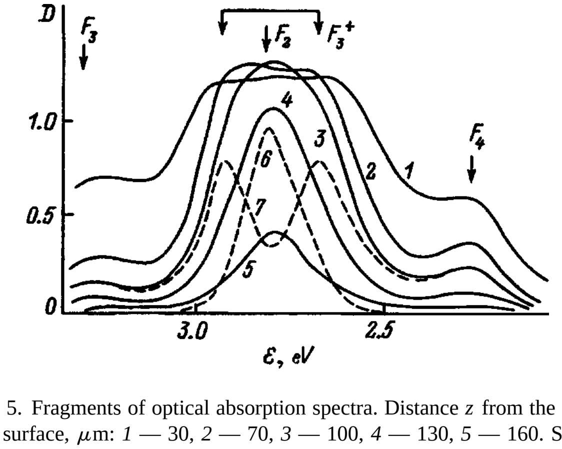

FIG. 5. Fragments of optical absorption spectra. Distance $z$ from the irradiated surface, $\mu$m: *1* — 30, *2* — 70, *3* — 100, *4* — 130, *5* — 160. Spectra *6* and *7* show the composition of spectral band *3*.

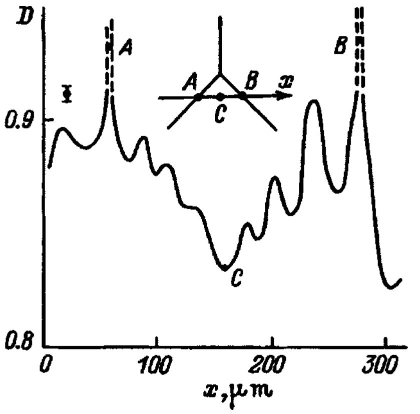

FIG. 6. Distribution of optical density ($\lambda = 445$ nm) in LiF irradiated by a high-current electron beam with surface energy density of 0.16 J/cm$^2$ ($J_{\max} = 40$ A/cm$^2$) measured in the $x$ direction (see the diagram of the measurement in the inset).

as we approach the channel the relative contribution to the absorption from 2.8 eV $F_2$ centers decreases, while that from $F_3^+$ centers increases. In this case we also observe an increase in the optical density at the maxima of the $F_3$ and $F_4$ bands. It should be noted that this spatial modulation of the absorption for 2.8 eV is typical of a thin crystal layer parallel to the irradiated surface when the former contains a breakdown channel track.

These results indicate that high-current electron beams with energy densities above threshold create distributions of $F$-aggregate centers over the irradiated volume that are significantly nonuniform, and that each type of center has its own characteristic dependence $D(\lambda,x,z)$. There are at least two fundamental reasons for this spatial nonuniformity. The first is the fact that the specific volume density of absorbed energy at a fixed coordinate $z$ is a function of time (Fig. 2), which accordingly leads to differing space–time rates of generation of electrons and holes in the sample. Consequently, the efficiency of the reaction that creates an $F_2^+$ center via capture of a hole by a $F_2$ center, like the reaction that creates an anion vacancy via the interaction of a $F$ center with a hole, will depend on the coordinate $z$ and the number of prior irradiation pulses for a crystal irradiated more than once, or, more precisely, on the concentrations of $F$ and $F_2$ centers created. The efficiency of the reaction that converts an $F_2$ center into an $F_2^+$ center via capture of a self-trapped hole (a $Y_k$ center) at temperatures above the detrapping temperature of the latter is also a function of coordinates, since the relative fraction of self-trapped holes present after an excitation pulse depends on the fraction of band holes that have reacted with $F$ and $F_2$ centers, or, more precisely, on the space- and time-dependent rate of generation of holes and the spatial distribution of electronic defects created previously. Qualitatively similar processes can probably be expected when more complex electronic centers are created, e.g., $F_3^+$, $F_3$, $F_4$ centers, etc.

A second structural element that participates in the conversion of electron centers is the anion vacancy. Therefore, it is important to understand the processes that generate vacancies and their spatial distribution. Anion vacancies are created in the stopping region of the beam electrons when $F$ centers capture holes, a process similar to that observed in KBr.[26] An additional channel that leads to generation of anion vacancies is the creation of these vacancies by dynamic stresses caused by the pulsed heating of the stopping zone by the electrons. These stresses can reach values of $10^7$ N/m$^2$ (Ref. 27). According to the data of Ref. 28, deforming stresses of this magnitude can lead to anion vacancy concentrations as large as $10^{18}$ cm$^{-3}$, and the defect density increases with increasing loading rate. It is logical to assume that the high concentration of anion vacancies created by dynamic loading should also participate in the conversion of electronic centers.

The second reason for spatial nonuniformity of the cumulative effects of irradiation is the development of multichannel electric breakdown, or, more precisely, of the following associated factors: a spatially nonuniform electric field, dynamic stresses, a thermal field caused by the dissipation of energy released in the breakdown channel, and electromagnetic radiation in the visible and IR regions. It is worth noting that each of these factors has its own intrinsic time dependence, so that it is difficult to determine the efficiency with which each of the factors affects the conversion of electronic centers. However, it is only the near-channel volume that is subjected to these dynamic stresses and temperatures, which lead to both increased concentrations of anion vacancies and increased rates of diffusion processes, so that this complexification of the structure of the electronic centers evolves in a well-defined way as the probe beam approaches the breakdown channel. Another factor to take into account is the increasingly complicated motion of anion vacancies due to the electric field, whose intensity in this region is a maximum (Fig. 3).

The processes that spatially modulate the concentrations of complex electronic defects as a breakdown channel is traversed (Fig. 6) are considerably more complex. The authors of Ref. 29 established that periodic damage structures (PDS) form in irradiated ionic crystals when the energy density of the high-current electron beam reaches a certain threshold value. In Ref. 30 it was shown that the generation of a PDS takes place synchronously with the time at which the high-current electron beam acts and is probably due to excitation of damaging stresses as the electromagnetic radiation generated by self-oscillating plasmas in the breakdown channel is absorbed. The agreement between the maxima in the modulated distribution of $F$-aggregate defects and the periodicity of alternating zones of damage in the PDS allows us to conclude that the modulated release of energy is due to the same mechanism in the two cases. This conclusion is verified by the fact that as the energy density of the high-current electron beam increases, the degree of modulation of the optical absorption increases too.

Finally, we must point out that the development of a breakdown channel in LiF in a plane parallel to the irradiated surface, due to the action of the strong electric field, must alter the conditions for stopping of the electron beam and consequently must change the energy release profile.

## CONCLUSION

Our experimental data and calculations allow us to clarify the primary reasons for nonuniform release of energy when a high-current electron beam acts on a high-resistance material.

When the energy density of the high-current electron beam is below the threshold for initiating multichannel breakdown, the primary cause of the nonuniform release of energy is the time dependence of the current and spectrum of the electron beam. U nder multiple irradiations this leads to spatial nonuniformity of the efficiency of the secondary reactions that convert electronic centers into complexes, and consequently to a complicated dependence of $D(\lambda,z)$.

At energy (current) densities of the electron beam that exceed the threshold for multichannel breakdown, the breakdown factors act to further distort $D(\lambda,x,z)$ by causing more complicated $F$-aggregate defects to accumulate as the damage channel is approached. In our view, the observed periodic modulation of $D(\lambda,z)$ comes about through excitation of characteristic lattice vibrations due to absorption of the electromagnetic radiation, one source of which is the self-oscillations of the electron–hole plasma in the breakdown channel. A more thorough analysis of this mechanism for energy release will be published as a separate article. Here we note only that as the energy density (current) of a high-current electron beam increases, the modulated release of energy leads to formation of periodic damage structures in LiF, KI, NaCl, and KCl,[29,30] implying that this mechanism for energy release is common to a number of alkali-halide crystals and is connected with microchannel breakdown. The low threshold for multichannel breakdown in ionic crystals is due to the favorable conditions for initiating thermionic and field emission: a high concentration of nonequilibrium electrons in the near-surface region of the crystal and the low surface potential barrier for ejection of nonequilibrium electrons.[18]

[1)] The results shown in Figs. 1, 2, and 3 were calculated using $J_{\max}=100$ A/cm$^2$ ($Q_{\mathrm{inj}}=1.2\times10^{-6}$ C/cm$^2$).

---

[1] R. B. Oswald, IEEE Trans. Nucl. Sci. **NS-13**(6), 63 (1966).
[2] D. I. Vaĭsburd (Ed.), *High-Energy Solid-State Electronics* [in Russian], Nauka, Novosibirsk (1982), 227 pp.
[3] A. K. Pikoev, A. S. Kabakichi, I. E. Makarov *et al.*, *Pulsed Radiolysis and its Applications* [in Russian], Atomizdat, Moscow (1980), 279 pp.
[4] E. D. Aluker, V. V. Gavrilov, R. G. Daĭch *et al.*, *Fast Radiation-Stimulated Processes in Alkali Halide Crystals* [in Russian], Zinatne, Riga (1987), 183 pp.
[5] V. I. Oleshko and V. F. Shtan'ko, Zh. Tekh. Fiz. **57**(9), 1857 (1987) [Sov. Phys. Tech. Phys. **32**, 1114 (1987)].
[6] V. F. Shtan'ko, V. I. Oleshko, and V. N. Inyakin, Fiz. Khim. Obrab. Mater. No. **6**, pp. 11–13 (1988).
[7] V. F. Shtan'ko and V. I. Oleshko, Zh. Tekh. Fiz. **59**, 99 (1989) [Sov. Phys. Tech. Phys. **34**, 312 (1989)].
[8] V. I. Oleshko and V. F. Shtan'ko, Fiz. Tverd. Tela (Leningrad) **29**, 320 (1987) [Sov. Phys. Solid State **29**, 182 (1987)].
[9] A. F. Akkerman, Yu. M. Nikitushev,and V. A. Botvin, *Solving Problems of Fast Electron Transport in Matter by the Monte Carlo Method* [in Russian], Nauka, Alma-Ata (1972), 163 pp.
[10] R. M. Sternheimer, Phys. Rev. **145**, 247 (1966).
[11] V. F. Shtan'ko, V. M. Tolmachev, and V. G. Glybin, Dep. in VINITI 3452-B95 at All-Union Institute of Scientific and Technical Information, Moscow (1995), 35 pp.
[12] D. E. Potter, *Computational Physics* [Wiley, New York (1973); Mir, Moscow (1975), 392 pp.].
[13] A. L. Gurskiĭ, E. V. Lutsenko, and G. P. Yablonskiĭ, IF AN BSSR Preprint No. 607 [in Russian] (Institute of Physics, Academy of Sciences of the BSSR, Minsk, 1990), 47 pp.
[14] V. I. Oleshko and V. F. Shtan'ko, Zh. Tekh. Fiz. **60**(2), 185 (1990) [Sov. Phys. Tech. Phys. Sov. Phys. Tech. Phys. **35**, 248 (1990)].
[15] A. A. Vorob'ev, *Physical Properties of Ionic Crystalline Insulators* (Tomsk State Univ., Tomsk, 1960) [in Russian], 232 pp.
[16] A. A. Vorob'ev and G. A. Vorob'ev, *Electric Breakdown and Destruction of Solid Insulators* [in Russian], Vyssh. Shkola, Moscow (1966), 224 pp.
[17] J. Nahum, Phys. Rev. **158**, 814 (1967).
[18] E. D. Akuler, D. Yu. Lusis, and S. A. Chernov, *Electronic Excitation and Radioluminescence of Akali-Halide Crystals* (Zinatne, Riga, 1979) [in Russian], 252 pp.
[19] C. J. Delbecq, Z. Phys. **171**, 560 (1963).
[20] Y. Farge, M. Lambert, and R. Smoluchowski, Solid State Commun. **4**, 333 (1966).
[21] B. P. Aduyev and D. I. Vaĭsburd, Fiz. Tverd. Tela (Leningrad) **20**, 3739 (1978) [Sov. Phys. Solid State **20**, 2165 (1978)].
[22] B. P. Aduyev and D. I. Vaĭsburd, Fiz. Tverd. Tela (Leningrad) **23**, 1869 (1981) [Sov. Phys. Solid State **23**, 1093 (1981)].
[23] V. I. Oleshko and V. F. Shtan'ko, Zh. Tekh. Fiz. **56**, 1235 (1986) [Sov. Phys. Tech. Phys. **31**, 727 (1986)].
[24] M. B. Fok, Proc. Fiz. Inst. Akad. Nauk SSSR **59**, 3 (1972).
[25] M. Okada, K. Atobe, and M. Nakogawa, *International Conference on Defects in Insulating Crystals*, Gatlinburg, Tenn. (1977), p. 321.
[26] Y. Kondo, M. Hivai, and M. Ueta, J. Phys. Soc. Jpn. **151**, 33 (1972).
[27] V. F. Shtan'ko, V. I. Oleshko, and V. M. Tolmachev, Fiz. Khim. Obrab. Mater. **2**, 53 (1991).
[28] B. I. Smirnov, Fiz. Tverd. Tela (Leningrad) **33**, 2513 (1991) [Sov. Phys. Solid State **33**, 1419 (1991)].
[29] V. M. Lisitsyn, V. I. Oleshko, and V. F. Shtan'ko, Pis'ma Zh. Tekh. Fiz. **181**(24), 1478 (1985) [Sov. Tech. Phys. Lett. **11**, 609 (1985)].
[30] V. I. Oleshko, and V. F. Shtan'ko, Zh. Tekh. Fiz. **57**, 2401 (1987) [Sov. Phys. Tech. Phys. **32**, 1457 (1987)].

Translated by Frank J. Crowne

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Depth distribution of vacancies generated by irradiating a solid surface with a flux of accelerated ions

Yu. D. Kornyushkin

*St. Petersburg Institute of Precision Mechanics and Optics (Technical University), 197101 St. Petersburg, Russia*

(Submitted June 4, 1996; resubmitted November 21, 1997)
Zh. Tekh. Fiz. **68**, 60–65 (April 1998)

Kinetic transport theory is used to find analytical expressions for the absorbed doses of the primary-particle flux and primary-particle energy as functions of distance into a solid with finite or semi-infinite thickness when the surface of the latter is irradiated by a flux of accelerated ions (atoms) in the direction normal to the surface. The theory was compared with experiments in which solid silicon films with thicknesses 50, 100, and 400 nm were irradiated by a flux of accelerated boron atoms with energies from 10 to 20 keV. These expressions were used to calculate the depth distribution of vacancies generated in a solid whose surface is irradiated by a flux of accelerated ions. The method developed can be used to determine the depth distribution of vacancies created by fluxes of accelerated electrons, neutrons, or photons.

[S1063-7842(98)01104-0]

## INTRODUCTION

The study of changes in the physical properties of a solid induced by irradiation of its surface with a flux of accelerated ions is valuable for both scientific and applications-related reasons. For instance, ion implantation of alkali and alkaline-earth elements can strongly modify the emission properties of materials.[1] Reproducible physical changes in the properties of materials find widespread application in ion-beam lithography.[2] The problem of the first wall of a thermonuclear reactor can be addressed to some extent by studying the changes that occur in a solid within the active zone of the reactor.[3,4] The last decade has seen the beginning of intense study of implantation in polyatomic targets, at present consisting of the constituent basis of semiconductors and insulators.[5] Note that ion implantation can be used to create high-temperature superconductors,[6] and that irradiation of a film by accelerated ions can cause it to have magnetic properties.[7] All these effects serve to motivate the study of interactions between currents of accelerated ions and matter.

In this paper, fluxes of primary atoms in a solid are modeled by transport theory in order to determine the energy they transport in the forward and reverse directions. Techniques are developed for computing the absorbed doses of atoms and energy at various depths and for calculating the depth distribution of vacancies.

## THEORY

When the surface of a solid target is irradiated by a flux of accelerated ions in the 1–100 keV energy range, elastic and inelastic collisions take place between primary particles and atoms of the target material. Because the energy losses due to elastic processes greatly exceed losses due to inelastic processes, the classical approximation can be used to describe the scattering of primary particles in the target. In this range of energies, the binding energy of a target atom to neighboring atoms of the solid target is much smaller than the energy of the primary particles. Therefore, the collision of a primary particle with a target atom can be treated as a pairwise interaction.

Assume that a flux of accelerated atoms (ions) is incident on the surface of a free solid film in the direction normal to the surface. Let us focus on an infinitesimally small volume $dV$ with thickness $dx$ at a depth of $x$. Fluxes of primary particles will propagate within the volume $dV$ in the forward and backward directions, and will also be absorbed. In order to describe the propagation of these fluxes in the material, we use the following kinetic transport equation, which holds within the microvolume of matter $dV$ at a depth $x$ and contains only average values of physical quantities:[8,9]

$$-wf(x,\mu)+w_s\int_{-1}^{1}W_s(x,\mu'\to\mu)f(x,\mu',\mu)\,d\mu' = \mu\frac{\partial f(x,\mu)}{\partial x}, \tag{1}$$

where $x$ is the coordinate of the microvolume $dV$ in the direction normal to the surface. Here $\mu'=\cos\theta'$, $\mu=\cos\Theta$, where $\Theta'$ and $\Theta$ are, respectively, the angle of incidence of the primary-particle flux on the elementary volume $dV$ at a depth $x$ and the scattering angle; $f(x,\mu)$ is the distribution function of primary particles in the material; $w=w_s+w_\gamma$, $w$ is the total macroscopic interaction cross section; $w_s$ and $w_\gamma$ are the macroscopic scattering and absorption cross sections, which we write as $w=\lambda^{-1}$, $w_s=\lambda_s^{-1}$, $w_\gamma=\lambda_\gamma^{-1}$, where $\lambda$, $\lambda_s$, and $\lambda_\gamma$ are average mean free paths for total scattering, elastic scattering, and absorption; and $W_s(x,\mu'\to\mu)$ is the scattering indicatrix, which is used to include anisotropy in the scattering of the primary particle flux within the solid:

$$W_s=\frac{dP_s(x,\Theta',\Theta)}{v\,d\Theta}, \tag{2}$$

where $dp_s(x,\Theta',\Theta)$ is the probability of a primary particle undergoing elastic scattering from the angular interval $d\Theta'$ to the interval $d\Theta$, and $v$ is the average velocity of the primary particle at a depth $x$.

Since the energy loss during atomic collisions is due primarily to elastic rather than inelastic processes in the energy range of interest,[10,11] we have $\lambda_s \ll \lambda_\gamma$. Consequently, $w \approx w_s$. Moreover, it is clear from general considerations that the average mean free path will be different at different depths $x$ in the target material. Therefore, to first approximation we can write

$$w = w_0 x, \tag{3}$$

where

$$w_0 - n_a \sigma,$$

$n_a$ is the concentration of target atoms, and $\sigma$ is the total microscopic cross section for the interaction of a primary particle with an atom of the target material. The method for computing this quantity is described in Ref. 12.

If we assume that the isotropic properties of the material do not depend on depth (i.e., $W_s(x,\mu' \to \mu)$ is a constant), then the following distribution function satisfies the kinetic transport equation plus its boundary conditions[11]:

$$f(x,\mu) = \frac{(1-r_0)d_0}{2v^2 B_1 \cosh(c_0)} \left( \frac{\exp(c_0(1-x^2/h^2))}{w_a - \mu} - \frac{\exp(-c_0(1-x^2/h^2))}{\omega_a + \mu} \right) F,$$

where $c_0 = a_0 h^2/2$, $w_a = w_0/a_0$, and

$$B_1 = w_a \ln \frac{w_a+1}{w_a-1} - 2;$$

$F = A_0 W_s + A_1$; $a_0$, $A_0$, $A_1$ are constants; $r_0$ is the integrated backscattering coefficient of the flux of primary particles from a free layer of thickness $h$; $\gamma_0$ is the flux density of primary atoms at the surface of the solid in the normal direction; and $h$ is the thickness of the solid layer.

The constant $a_0$ can be written in terms of the maximum longitudinal range of primary particles in the target material $l_p$, whose value we define as the thickness of a layer through which 0.01 of the initial primary particle flux passes. Then, starting from this definition, one can evaluate the maximum longitudinal range from the expression

$$l_p = \lim_{x \to l_p} (J_1(x)/J_0(x)), \tag{4}$$

where

$$J_0(x) = \int_0^x \frac{dn}{dx} dx, \quad J_1(x) = \int_0^x x \frac{dn}{dx} dx,$$

$n$ is the number of accelerated primary particles per unit volume of material at a depth $x$.

According to the physical meaning of the distribution function, one can write

$$f(x,\mu) = \frac{dn}{v\,d\mu}.$$

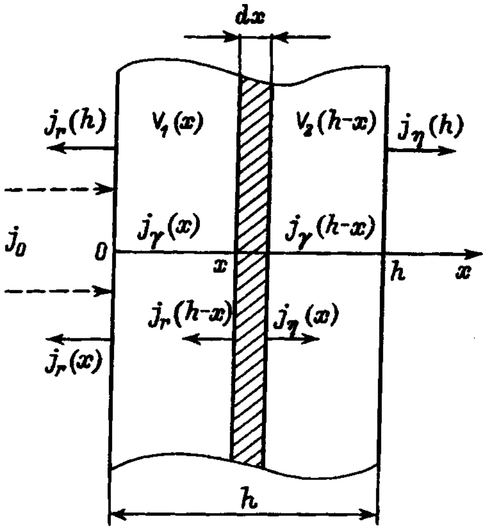

FIG. 1. Nominal sketch of primary particle fluxes within a free layer and outside it.

Consequently, $dn = vf(x,\mu)d\mu$. Substituting the distribution function into Eq. (4) gives $a_0 \approx 3l_p^{-2}$. It follows from experiment that

$$a_0 \approx 1.455 R_p^{-2}, \tag{5}$$

and therefore $l_p = 1.436 R_p$. The value of $R_p$ can be computed from the expression given in Ref. 13:

$$R_p = c_1 \frac{A_2 (Z_1^{2/3} + Z_2^{2/3})^{1/2}}{Z_1 Z_2 \rho_2} E_p, \tag{6}$$

where $E_p \geq 10$ keV; $c_1 = 137.4(a_2 + b_2 \ln(A_2/A_1))$; $A_2 > 2A_1$; $a_2 = 0.6366$; $b_2 = 0.0611$; $Z_1$ and $Z_2$ are the atomic numbers of the elements of the primary atoms and the target; $A_1$ and $A_2$ are the atomic weights of the primary atoms and target atoms; $\rho_2$ is the density of the target material; and $E_p$ is the kinetic energy of the primary particles.

Let the elementary volume $dV$ for which the kinetic transport Eq. (1) was written be at a depth $x$. Then the total volume $V$ can be divided up into two auxiliary volumes $V_1(x)$ and $V_2(h-x)$ (Fig. 1). Irradiation of the surface by a particle flux $j_0$ gives rise to forward $f_\eta(x)$ and backward $j_r(h-x)$ fluxes in each of these auxiliary volumes and also to absorption of the primary particles as a result of multiple collisions with target atoms.

Writing a balance equation for the fluxes of primary particles in the volume $V_2(h-x)$ gives

$$j_\eta(x) = j_r(h-x) + h_\gamma(h-x) + h_\eta(h), \tag{7}$$

where $j_r(h-x)$ is the flux density of backscattered primary particles from a layer of thickness $h-x$ at a depth $x$; $j_\eta(x)$ is the flux density of primary particles that have passed through a layer of thickness $x$ and have therefore penetrated into the volume $V_2(h-x)$; $j_\gamma(h-x)$ is the flux density of primary particles absorbed in the volume $V_2(h-x)$; and $j_\eta(h)$ is the flux density of primary particles leaving the free film of thickness $h$.

When no sources of primary particles are present in an elementary volume $dV$ there exist two physically indistinguishable fluxes of primary particles propagating in opposite directions: $j_\eta(x)$ and $j_r(h-x)$. Therefore, the resulting flux of primary particles in the volume $dV$ will be

$$j(x)=j_\eta(x)-j_r(h-x). \tag{8}$$

Then after substituting $j_\eta(x)$ from Eq. (7) into Eq. (8), we obtain (Fig. 1)

$$j(x)=j_\gamma(h-x)+j_\eta(h). \tag{9}$$

Equation (9) can be written in terms of transmission and absorption coefficients:

$$\eta(x)=\gamma(h-x)+\eta(h), \tag{10}$$

where $\eta(x)=j(x)/j_0$, $\gamma(h-x)=j_\gamma(h-x)/j_0$, and $\eta(h)=\gamma_\eta(h)/j_0$.

As follows from Ref. 11,

$$\eta(x)=\frac{1-r_0}{\cosh(c_0)}\cosh(c_0(1-x^2/h^2)), \tag{11}$$

where $r_0$ is the integrated backscattering coefficient from a free layer of thickness $h$.

The integrated backscattering coefficient from a layer of thickness $h-x$ can be computed from the expression

$$r(h-x)=\frac{1-r_0}{\cosh\,(c_0)}\,\frac{B_2}{B_1}\sinh\,(c_0(1-x^2/h^2)),$$

$$r_0=\left(1-\frac{B_1}{B_2}\coth\,(c_0)\right)^{-1}. \tag{12}$$

In general the coefficients $B_1$ and $B_2$ can be computed from the expressions given in Ref. 11:

$$B_1=\int_0^{2\pi}d\varphi\int_0^{\pi}\frac{F(\mu')\sin 2\Theta' d\Theta'}{w_a-\mu},$$

$$B_2=\int_0^{2\pi}d\varphi\int_{\pi/2}^{\pi}\frac{F(\mu')\sin 2\Theta' d\Theta'}{w_a-\mu}.$$

If $W_s(x,\mu'\to\mu)$ in Eq. (2) is independent of the angles $\Theta'$ and $\Theta$, then calculations show that

$$B_1=w_a\ln\frac{w_a+1}{w_a-1}-2,\quad B_2=w_a\ln\frac{w_a}{w_a-1}-1.$$

We now can use Eqs. (8) and (12), according to Eq. (7), to calculate the flux of primary particles at an arbitrary depth $x$ that propagate only in the forward direction:

$$\eta_1(x)=\frac{1-r_0}{\cosh(c_0)}(\cosh(1-x^2/h^2)) + \frac{B_2}{B_1}\sinh(C_0(1-x^2/h^2)). \tag{13}$$

After substituting Eq. (11) into Eq. (10), we find that the absorption coefficient for the flux of primary particles in a layer of thickness $h-x$ is

$$\gamma_2(h-x)=\frac{1-r_0}{\cosh(c_0)}(\cosh(c_0(1-x^2/h^2))-1), \tag{14}$$

while the absorption coefficient in a layer of thickness $x$ is

$$\gamma_1(x)=\gamma_2(h)-\gamma_2(h-x),$$

where $\gamma_2(h)=\gamma_2(h-x)|_{x=0}$.

Consequently,

$$D(x)=\gamma_1(x)=\frac{1-r_0}{\cosh(c_0)} \times(\cosh(c_0)-\cosh(c_0(1-x^2/h^2))). \tag{15}$$

Equation (15) allows us to compute the magnitude of the absorbed flux of primary particles in a layer of thickness $x$, i.e., to calculate the absorbed primary particle dose. If, however, the average energy of the primary particles in a layer of thickness $x$ is known, then the absorbed energy dose will be

$$D(x,E)=\gamma_1(x)\Delta E, \tag{16}$$

where $\Delta E$ is the average energy loss from the primary particles at a depth $x$.

The average energy loss from the primary particles can be computed using the expressions given in Ref. 2, according to which the average relative energy losses for elastic $b_e$ and inelastic $b_i$ scattering at a depth $x$ are

$$b_e=b_1x,\quad b_i=b_2x,$$

where

$$b_1=\frac{D_1Z_1Z_2A_1\rho_2E_p^{-1}}{(Z_1^{2/3}+Z_2^{2/3})(1+A_2^{-1})},$$

$$b_2=\frac{D_2(Z_1+Z_2)\rho_2}{A_1^{1/2}A_2E_p^{1/2}}, \tag{17}$$

and $D_1=167.36$, $D_2=6240$.

Then the total relative energy loss will be

$$b=b_e+b_i, \tag{18}$$

and the average energy loss from the primary particles at a depth $x$ is

$$\Delta E=bE_p. \tag{19}$$

Expressions (13) and (19) allow us to calculate the average relative energy transported by the flux of primary particles through an elementary volume $dV$ at a depth $x$ in only the forward and only the reverse direction:

$$\eta_1(x,E)=(1-bx)\eta_1(x),$$

$$r(x,E)=(1-bx)r(x),\quad x\leqslant h\leqslant l_p. \tag{20}$$

Equation (16) can be quite valuable in applications, since it simplifies the calculation of absorbed doses of ionizing radiation in humans and animals, which is extremely complicated to do under experimental conditions.

From Eq. (15) it follows that the depth profile of the absorbed radiation, when Eq. (5) is taken into account, should have the form

$$G(x)=\frac{d\gamma_1(x)}{dx}=\frac{(1-r_0)a_0x}{\cosh(c_0)}\sinh(C_0(1-x^2/h^2)). \tag{21}$$

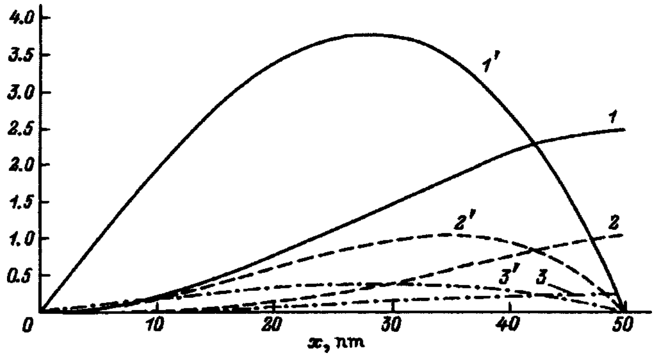

FIG. 2. $h = 50$ nm, $E_p = 10$ keV: *1* — dose of absorbed particles $20 \cdot D(x)$; *1'* — distribution of absorbed particles with respect to depth $x$, $10^4 \cdot G(x)$; *2* — dose of absorbed energy, $20 \cdot D(x,E)$; *2'* — depth distribution of dose of absorbed energy, $10^4 \cdot G(x,E)$; $E_p = 20$ keV: *3* — dose of absorbed particles $20 \cdot D(x)$; *3'* — depth distribution of dose of absorbed particles, $10^4 \cdot G(x)$.

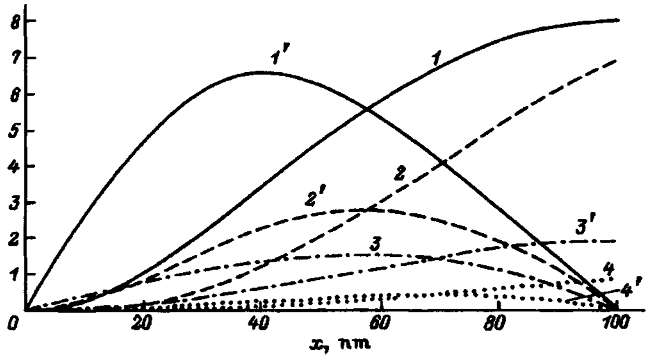

FIG. 3. $h = 100$ nm; *1, 1', 2, 2'* — the same as in Fig. 2, with $E_p = 10$ keV; $E_p = 20$ keV: *3* — dose of absorbed energy, $20 \cdot D(x,E)$; *3'* — depth distribution of dose of absorbed energy, $10^4 \cdot G(x,E)$; *4* — dose of absorbed energy, $20 \cdot D(x,E)$; *4'* — depth distribution of dose of absorbed energy, $10^4 \cdot G(x,E)$.

The depth profile of the absorbed primary-particle energy (the absorbed energy dose) according to Eqs. (16)–(21) can be written in the form

$$G(x,E) = \frac{dD(x,E)}{dx} = (b_1 + b_2)\gamma_1(x) + b\frac{d\gamma_1(x)}{dx}. \quad (22)$$

## RESULTS AND DISCUSSION

In Ref. 12 a method was proposed to calculate the total microscopic cross section for interaction of atoms in a flux of primary particles with atoms of a target material. The results of these calculations for the interaction of boron atoms with silicon atoms in the energy range under discussion are found to be in good agreement with the results of Kumakhov *et al.*,[14] which in turn confirms the correctness of the approach used to estimate the magnitude of the effective microscopic inteaction cross section.

Equations (15), (16), (21), and (22) allow us to calculate doses of absorbed primary particle flux and the energy transferred by the primary particles, and also the distribution of these doses with respect to depth for various thicknesses of the solid film. These calculations were done for Si layers irradiated by a flux of accelerated boron ions. To make these results easier to visualize, we used layers with thicknesses 50, 100, and 400 nm and accelerated the ions to energies of 10 and 20 keV. We found that this choice of initial values for the layer thickness and primary particle energies enabled us to arrive at several important qualitative conclusions.

Figure 2 shows the dose functions $D(x)$, $D(x,E)$ and their depth distributions $G(x)$, $G(x,E)$ in a layer of thickness 50 nm for boron ion energies of 10 and 20 keV. Because in this case the maximum longitudinal range $l_p > 50$ nm, we can assume that a significant number of primary particles leave the free layer. From Fig. 2 it is clear that the depth profiles of the particle and energy doses do not have their maxima at the same position: the maximum of the absorbed energy dose profile is shifted farther into the bulk. Moreover, the absorbed dose decreases strongly as the energy increases. Figure 3 shows the same dependences as Fig. 2, but for a thickness of 100 nm, while Fig. 4 shows them for a thickness of 400 nm.

Figure 4 shows the functions $D(x)$, $D(x,E)$, $G(x)$, and $G(x,E)$ for a free film of thickness 400 nm. For primary particle energies $E_p \leqslant 20$ keV the thickness of such a layer may be treated as infinitely large, since no flux of primary particles passes through it in the forward direction. Comparing the functions $G(x)$ and $G(x,E)$ in Fig. 4 with the corresponding functions in Figs. 2 and 3, we note that in Fig. 4 they monotonically approach the abscissa. This difference in the shapes of the curves allows us to establish qualitatively when the thickness of a film exceeds the average maximum mean free path of a primary particle, and when it is smaller. From Fig. 4 it is also clear that the position of the distribution maxima with respect to depth for the absorbed dose of primary particles $G(x)$ and the absorbed energy dose $G(x,E)$ do not coincide. Therefore, e.g., in the implantation of boron ions in silicon one should use the function $G(x)$, and to obtain the depth profile of the energy absorbed by the material one should use the function $G(x,E)$. This approach must be used in estimating radiation damage to a human or animal.

Thus, when estimating the absorbed dose using equations (15), (16), (21), (22), we must take into account the

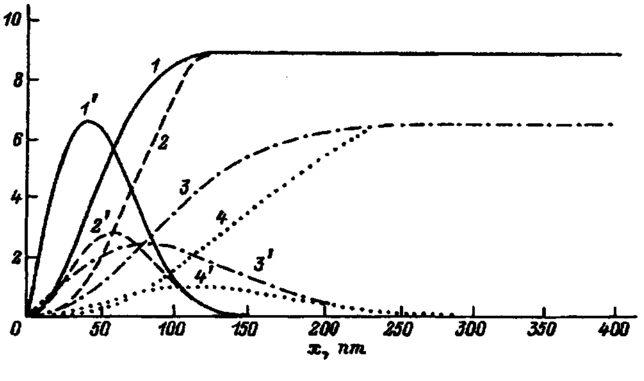

FIG. 4. $h = 400$ nm: $E_p = 10$ keV: *1, 1', 2, 2'* — the same as in Fig. 2; $E_p = 20$ keV; *3* — dose of absorbed particles, $20 \cdot D(x)$; *3'* — depth distribution of dose of absorbed particles, $10^4 \cdot D(x)$; *4* — dose of absorbed energy, $20 \cdot D(x,E)$; *4'* — depth distribution of dose of absorbed energy, $10^4 \cdot G(x,E)$.

physical and chemical features of the primary particles. If the primary particles are chemically active in an organism, then the primary effect of these particles on that organism will be determined by values of $D(x)$, $G(x)$. If, however, the energy of the primary particles is high enough that their absorption leads to considerable change in the tissue, then the most important thing to do is to determine the values of $D(x,E)$ and $G(x,E)$. By using the expressions derived in this paper, such an estimate can be made rather simply.

Thus, the method developed here allows us to calculate with reasonable accuracy both the absorbed dose and the distribution of absorbed primary particles within a layer of thickness $x$. The distribution that results when a flux of boron ions at energies of 10, 15, and 20 keV irradiates the surface of a free silicon film with a thickness of 100 nm in the direction normal to the surface was calculated in Ref. 12. The results given there are in good agreement with the experiments of Ref. 15 and completely correspond with the results of the present paper.

In general, the primary particles lose energy in elastic and inelastic collisions, and also by creating vacancies. The process of implantation is accompanied both by amorphization of the target crystal lattice and by the production of vacancies in it.[16] For primary particle atoms with energies above 10 keV, the region where the largest absorption of primary particle flux takes place is probably the region where vacancies are generated. The region of intense generation of vacancies is at a depth where the most intense collisions of primary particles with atoms of the material takes place. Therefore, the distribution of vacancies in matter, and consequently the distribution with respect to depth of the amorphized layer, should coincide with the depth distribution of primary atomic particles absorbed in the material. In Fig. 4 we show the depth distributions of the absorbed primary particles and the absorbed energy dose at normal incidence. In Ref. 16 Motooka and Holland studied the effect of implantation of $\langle 110 \rangle$ 100 keV single-crystal silicon with 100 keV Si ions and 175 keV Ge, As ions. It was established experimentally that an amorphous layer and vacancies appear in the sample.

For a more detailed theoretical analysis, the implantation process was simulated by the Monte Carlo method, for which the TRIM code was used.[17] The trajectories of ten thousand ions were followed. The threshold energy change corresponding to the appearance of a vacancy in a cascade was chosen to equal 13 eV, for a binding energy between lattice atoms of 4.5 eV. Figure 5 shows the depth distribution of vacancies in a silicon layer whose surface is irradiated by a flux of accelerated arsenic ions calculated in Ref. 16, along with the depth distribution of boron atoms obtained in the present paper. A comparison of these profiles shows that they are in rather good agreement. In this way we confirm the correlation between the depth distributions of absorbed atoms in a material and the depth distribution of vacancies created by collisions.

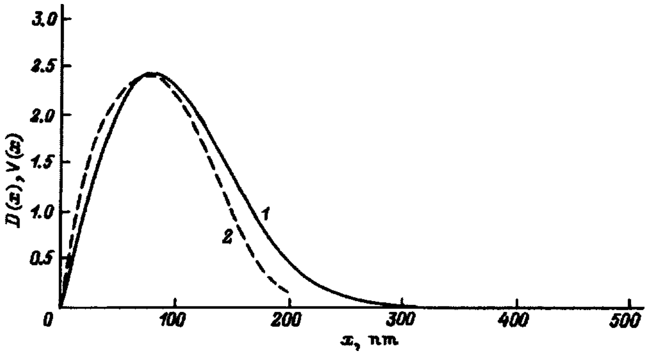

FIG. 5. (*1*) Depth distribution of dose $D(x)$ of absorbed 20 keV boron atoms in a free silicon layer with thickness 400 nm and (*2*) depth distribution of vacancies $V(x)$ arising in silicon when the surface is irradiated with 175 keV arsenic atoms (Ref. 16). $D(x)$ is in relative units, and $V(x)$ is vacancies/ion in relative units.

Thus, the distribution of absorbed flux of primary atomic particles in a single-crystal material can be used to determine the depth distribution of the amorphized layer, its half-width, and the deposition depth. Conversely, the depth distribution of the amorphized layer can be used to determine the depth distribution of energy flux absorbed from the primary atoms.

[1] A. Kh. Kasymov, *Surface Properties of Solids Doped by Ion Bombardment* [in Russian], Fan, Tashkent (1987), 130 pp.

[2] I. Brodie and J. J. Muray, *The Physics of Microfabrication,* Plenum Press, New York, 1982 [Russian trans. Mir, Moscow, 1985], 496 pp.

[3] R. A. Lagby, J. Bohdansky, W. Eckstein *et al.*, Nucl. Fusion: Special Issue (1984), 21 pp.

[4] *Atomic and Plasma-Material Interaction Data for Fusion*, Nucl. Fusion Suppl. **1**, 1 (1991).

[5] D. Fink, J. P. Biersack, M. Behar *et al.*, Appl. Phys. **58**, 668 (1985).

[6] Wang Ke-Ming, Liu Xi-Ju, Wang Ji-Hua *et al.*, , J. Phys. D **21**, 1624 (1988).

[7] M. S. Dresselhaus and A. Lusnikow, Synth. Met. **23**, 401 (1988).

[8] Yu. D. Kornyushkin, Fiz. Tverd. Tela (Leningrad) **20**, 1175 (1978) [Sov. Phys. Solid State **20**, 676 (1978)].

[9] Yu. D. Kornyushkin, Poverkhnost', No. 12, pp. 13–18 (1992).

[10] É. S. Parilis, N. Yu. Turaev, F. F. Umarov, and S. A. Nizhnaya, *Theory of Scattering of Medium-Energy Atoms by a Solid Surface* [in Russian], Fan, Tashkent (1987), 212 pp.

[11] V. A. Kvilidze and S. S. Krasil'nikov, *Introduction to the Physics of Atomic Collisions* [in Russian], Moscow State Univ. Publ., Moscow (1985), 224 pp.

[12] Yu. D. Kornyushkin, Poverkhnost', No. 7, pp. 40–96 (1992).

[13] M. A. Kumakhov and F. F. Komarov, *Energy Losses and Ranges of Ions in Solids* [in Russian] (Minsk, 1979), 320 pp.

[14] R. Ito, T.Tabata, T. Shirai, and R. A. Phaneuf, JAERI-M rep., Tokyo (1993), 93 pp.

[15] B. N. Mukashev, V. V. Smirnov, Z. Kalbitser, and M. Veǐzer, Poverkhnost', No. 3, p. 72–78 (1990) [in Russian].

[16] T. Motooka and O. W. Holland, Appl. Phys. Lett. **61**, 3005 (1992).

[17] J. P. Biersack, Nucl. Instrum. Methods **174**, 257 (1980).

Translated by Frank J. Crowne

# Effect of thermal annealing on the magnetic properties of thin films of Co–Pd alloys

S. A. Gusev, Yu. N. Nozdrin, D. B. Rozenshteĭn, and A. E. Tselev

*Institute of Microstructure Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia*

(Submitted December 4, 1996)
Zh. Tekh. Fiz. **68**, 66–70 (April 1998)

The results of studies of the effect of thermal annealing on the magnetic properties of thin films of Co–Pd alloys are described, along with a method for obtaining easy-axis perpendicular anisotropy in these materials. The method consists of depositing layers of the alloy on a substrate whose thermal expansion coefficient is considerably smaller than that of the film, and then annealing it in a vacuum. This method is used to prepare samples with rectangular hysteresis loops for magnetization perpendicular to the plane of the film, coercive forces of 750 Oe, and Kerr rotation angles of 0.21°. Also presented are the results of experiments on thermomagnetic recording on the samples. © *1998 American Institute of Physics.* [S1063-7842(98)01204-5]

In recent years the multilayer Co/Pd and Co/Pt structures have been studied intensely. Their large easy-axis perpendicular anisotropy, large Kerr rotation angle in the spectral region around 400 nm, and high corrosion resistance make them promising media for the magnetooptic devices to be used in the next generation of information storage systems. The perpendicular anisotropy of these multilayer structures is due to properties of the boundaries between the Co and noble metal layers, i.e., it is a surface effect. A number of theoretical and experimental papers (see, e.g., Refs. 1 and 2) have shown conclusively that the value of the perpendicular anisotropy depends strongly on the sharpness of the boundary between the layers, and decreases as the boundaries become smeared out. This fact makes the technology of fabriacting multilayer Co/Pd and Co/Pt structures with the desired properties rather complicated. However, a number of papers in which the magnetic and magnetooptic properties of thin films of Co–Pd and Co–Pt alloys prepared by various methods were investigated have reported that easy-axis perpendicular magnetic anisotropy can be obtained in these alloys. In particular, films of Co–Pd alloy with perpendicular anisotropy were obtained by the authors of Refs. 3–5 using magnetron sputtering and electrolytic deposition, and by the authors of Ref. 6 using molecular-beam epitaxy.

In this paper we investigate the effect of thermal annealing on the magnetic properties of Co–Pd alloy films, and develop a method for preparing films with perpendicular magnetic anisotropy that are suitable as magnetooptic recording media.

## THE POSSIBILITY OF OBTAINING EASY-AXIS PERPENDICULAR ANISOTROPY BY THERMAL ANNEALING

The magnetic anisotropy constant of a thin film can be expressed as a sum

$$K_{\mathrm{eff}}=K_{\mathrm{mc}}+K_{\sigma}-2\pi M_s^2,$$

where $K_{\mathrm{mc}}$ is the contribution due to magnetocrystalline anisotropy, $K_{\sigma}$ is the contribution due to elastic stress in the film, and the third term describes shape anisotropy; $M_s$ is the saturation magnetization, and $K_{\sigma}=(3/2)\lambda\sigma$, where $\lambda$ is the magnetostriction constant of the film material and $\sigma$ is the value of the elastic stress.

In keeping with the customary conventions, a positive sign for $K_{\mathrm{eff}}$ corresponds to easy-axis perpendicular anisotropy. The disordered Co–Pd alloys have a face-centered cubic lattice, and the magnetocrystalline anisotropy in them is quite small. However, Co–Pd alloys also exhibit considerable magnetostriction. In alloys with Co concentrations of around 25 at. % the magnetostriction constant is negative and can be as large as $\lambda=-1.5\times10^{-4}$ (Ref. 7). Therefore, elastic stresses can have a large effect on the magnitude of the anisotropy in this alloy. In Ref. 3 Hashimoto *et al.* showed that in multilayer Co/Pd structures magnetostriction can make a large contribution to the perpendicular anisotropy. In Ref. 8 by the same authors, perpendicular magnetic anisotropy was observed in films of the alloy Co–Pd prepared by magnetron sputtering. Perpendicular anisotropy in these films occurs only for certain sputtering regimes, and was associated by Hashimoto *et al.* with tensile elastic stresses caused by peculiarities in the process of film growth. Essentially the same phenomena was noted by Tsunashima *et al.* in Ref. 9. According to estimates by the authors of the aforementioned papers, the elastic stresses in alloy films with perpendicular anisotropy are in order of magnitude equal to $10^9$ dyn/cm$^2$.

The same magnitude of tensile elastic stress can be obtained by annealing films deposited on a substrate with a thermal expansion coefficient considerably smaller than that of the film. In fact, keeping the film at a high temperature during the anneal leads to relaxation of the elastic stresses in it. However, if the cooling is not too slow, tensile elastic stresses will be stored in a film grown on a substrate with a smaller thermal expansion coefficient. Let us assume that the thermal expansion coefficient of the film equals the thermal expansion coefficient of pure Pd ($\sim130\times10^{-7}$ 1/K). Then when such a film is grown on a substrate made of quartz glass (whose thermal expansion coefficient is $\sim5\times10^{-7}$ 1/K) and annealed at a temperature of 500°, its strain relative to the free state is 0.5% and the elastic stress

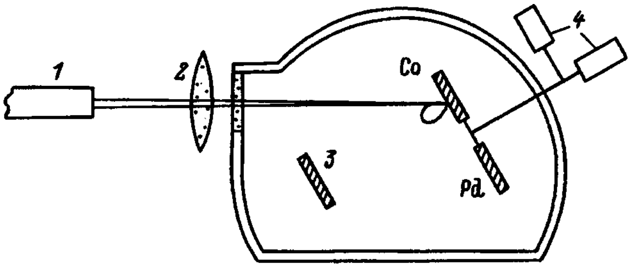

FIG. 1. Sketch of the laser deposition setup. *1* — Nd:YAG laser, *2* — focusing lens; *3* — substrate, *4* — stepper motor used to exchange targets and scan targets under the laser beam; the targets are Co and Pd.

can reach values of $10^{10}$ dyn/cm$^2$ (if we ignore the fact that elastic stresses are partially lowered by plastic deformation in the calculation), while the magnetic anisotropy constant, for example in an alloy film with a Co concentration of 30 at. %, is $2.2\times10^6$ erg/cm$^2$.

Thus, annealing of Co–Pd alloy films grown on substrates with considerably smaller thermal expansion coefficients can be a method for obtaining media with easy-axis perpendicular anisotropy. In this paper we establish that it is indeed possible to obtain films of the alloy Co–Pd with perpendicular magnetic anisotropy by this method, and investigate several magnetic and magnetooptic properties of such structures.

## SAMPLE PREPARATION

Co–Pd alloy films were prepared by pulsed laser deposition. A sketch of the deposition setup is shown in Fig. 1. In this apparatus we use a neodymium laser made of a yttrium orthoaluminate crystal with a wavelength of 1.06 $\mu$m (ILTI-207). For energies of 0.7 J per pulse, a power flux density of $\sim2\times10^9$ W/cm$^2$ is created at the target. Deposition takes place in vacuum at a pressure of $2\times10^{-6}$ Torr from two different targets made of Co and Pd illuminated one after another onto a substrate at room temperature. It is well known that the cloud of plasma evaporated by an intense laser beam in vacuum contains a certain fraction of particles with energies above 500 eV.[10] When such particles collide with the surface of a growing film, they are capable of strongly disrupting its top layer and burying themselves in the thickness of its bulk to a certain depth. Specially designed deposition experiments and investigations of small-angle diffraction by multilayer periodic Co/Pd structures with a period of around 100 Å have shown that the thickness of the mixed layer at the boundary between Co and Pd layers is roughly 10 Å. This fact dictates that as the alloy films are deposited, the number of pulses of the sputtering laser arriving at a target during one period be chosen such that the period of modulation of the film composition should not exceed 10 Å. Using this approach, it is expected that the film will consist of a practically uniform mixture of the two materials. The average composition of the film (concentration) is controlled by varying the relative numbers of laser pulses arriving at the Co and Pd targets. The magnetic properties of the film were studied using the polar and meridional Kerr effects at a wavelength of 630 nm. The composition of the films was determined by x-ray microanalysis.

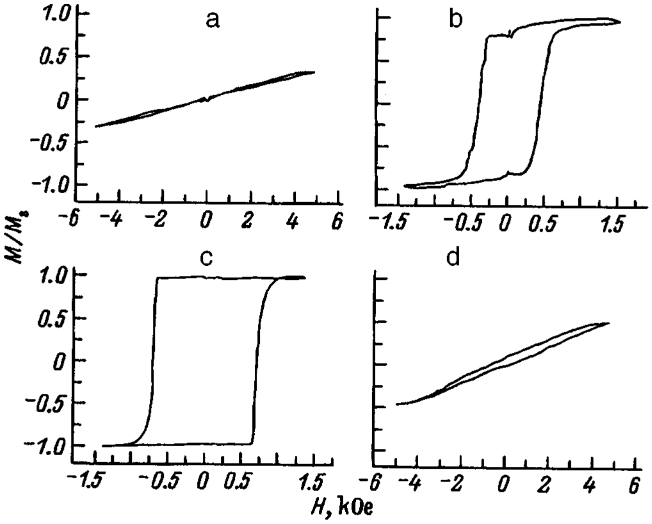

FIG. 2. Magnetization curves for one of the films.

The samples were annealed in a vacuum chamber with a residual gas pressure of no more than $5\times10^{-6}$ Torr, and cooled naturally in the vacuum. The time for heating to a temperature of 300 °C was 4 minutes, and was 2 minutes from 300 500 °C. Cooling from 500 to 300 °C took 5 minutes. The anneal time obviously should be no smaller than the time required for relaxation of elastic stresses at the anneal temperature. Experiments show that at a temperature of 200 °C the magnetic properties of the film cease to change noticeably for anneal times of more than two hours; at a temperature of 300 °C this happens after more than 1 hour, at a temperature of 400 °C it happens after 15 minutes, and at a temperature of 500 °C 4 minutes is sufficient for stability. The x-ray diffraction and electron-microscope studies established that as-deposited films consist of polycrystals with grain sizes of 100−150 Å without any texture. Anneal temperatures up to 400 °C change the film structure slightly. Annealing at temperatures higher than 500 °C causes the grain size to increases to 450 Å and the film acquires a texture with the (111) axis perpendicular to the plane of the film. Rocking curves around the (111) peak of the large-angle x-ray diffraction pattern have a width at half-height of 4−9°, indicating a high degree of texturization of the films.

## RESULTS OF EXPERIMENTS

Figure 2 shows how the magnetic anisotropy changes as a result of annealing for a film with thickness 150 Å and Co concentration of 25 at. % deposited on a substrate made of quartz glass. In Figs. 2a and 2b, curves for the magnetization are shown for as-deposited films and fields applied perpendicular and parallel to the plane of the sample respectively, obtained from the polar (a) and meridional (b) Kerr effect. Figures 2c and 2d show the same dependences for the same sample after annealing at a temperature of 500 °C for 5 minutes. It is clear that immediately after deposition the film has a strong anisotropy of the easy-plane type. Annealing leads

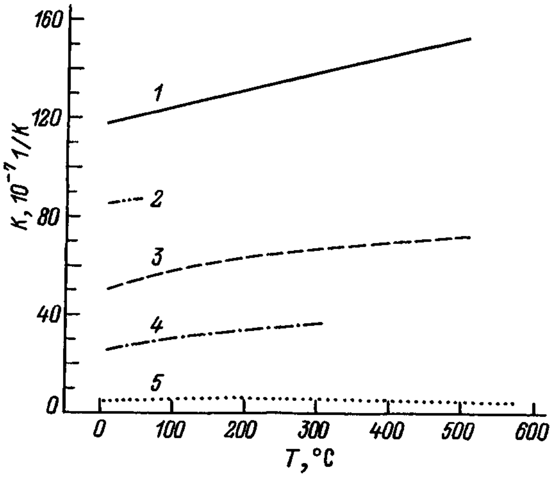

FIG. 3. Temperature dependence of the thermal expansion coefficient (K) of Pd and the substrate materials used in this paper. *1* — Pd, *2* — BF12-type optical glass; *3* — sapphire $Al_2O_3$, *4* — Si, *5* — KB-type quartz glass.

to a change in the sign of the anisotropy, while the hysteresis loop, which is taken in a field perpendicular to the film plane, becomes almost completely rectangular in form with a coercive force of 750 Oe. This type of change in the anisotropy is typical for samples deposited on quartz glass.

As we assumed, thes changes in the magnetic anisotropy of the film are due to the different thermal expansion coefficients of the film and substrate materials. In order to verify that it is the difference in thermal expansion coefficients of the film and substrate that is the decisive factor in changing the anisotropy of the film, we performed a number of experiments on the deposition and annealing of films on different substrates. Films with the same composition and thickness were deposited on substrates made of KB-type quartz glass, silicon, sapphire, and BF12-type optical glass. Figure 3 shows the dependence of the thermal expansion coefficient on temperature for these materials according to handbook data.[11] In order to prevent a chemical reaction between the silicon and the alloy, we used the method of vacuum-arc sputtering to deposit a buffer layer of amorphous diamond-like carbon with a thickness of 200 Å onto the silicon substrate. The results of these experiments are shown in Figs. 4, 5, and 6. Figure 4 shows how the shape of the hysteresis loop changes in applied magnetic fields perpendicular to the plane of the film, and the increase in anneal temperature for the example of a film on a substrate made of quartz glass. In order to measure the characteristics of the hysteresis loops perpendicular to the film plane, we chose the following fields (Fig. 4): $H_c$ is the coercive force, $H_n$ is the field at which domains nucleate in films that possess perpendicular anisotropy; and $H_s$ is the field at which films with easy-plane anisotropy saturate. It is obvious that the magnitudes of these fields are directly related to the value of the magnetic anisotropy for films prepared using the same technology. Figure 5 shows the functions $H_c$ and $H_n$ plotted versus the average thermal expansion coefficient of the substrate at two anneal temperatures; in Fig. 6 we show the dependence of the same quantities on anneal temperature for substrates made of quartz and optical glass.

From these plots it is clear that the magnitude of the change in anisotropy is uniquely determined on the one hand by the anneal temperature, and on the other hand by the thermal expansion coefficient of the substrate. The smaller the thermal expansion coefficient is (i.e., the larger the difference between the thermal expansion coefficient of the film and that of the substrate), the smaller is the temperature required to obtain the same change in anisotropy. Whereas starting from a certain anneal temperature (which depends on the thermal expansion coefficient of the substrate) further increases in the temperature do not lead to a change in the anisotropy. Such behavior of the anisotropy is probably explained by the fact that starting with a certain value of the

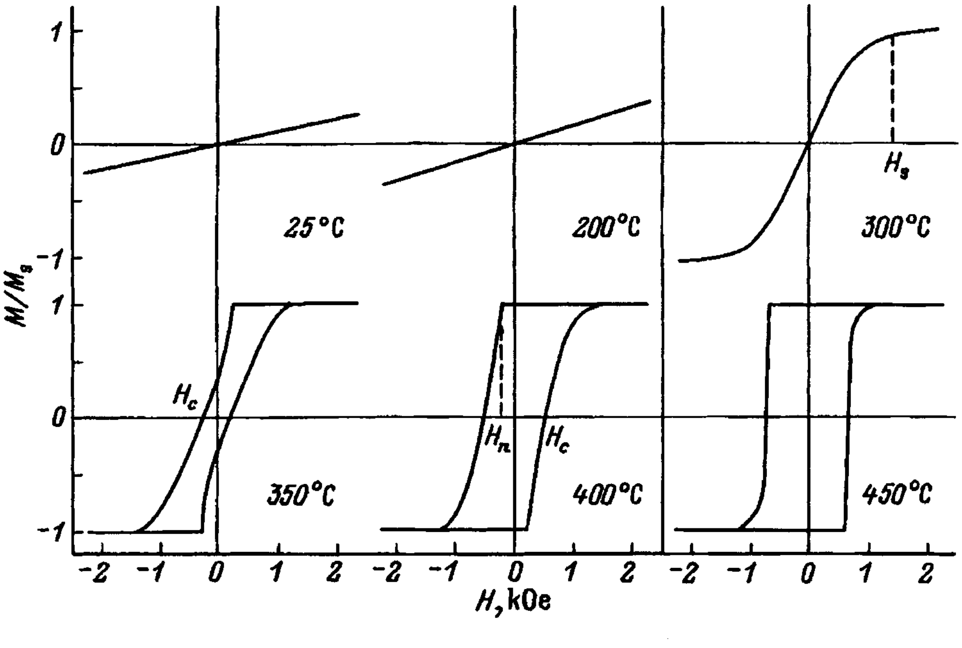

FIG. 4. Changes in the shape of hysteresis loops as the anneal temperature is increased.

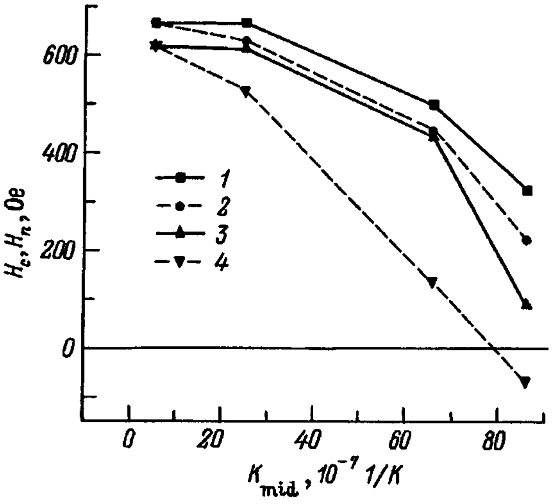

FIG. 5. Coercive force $H_c$ (*1,2*) and domain nucleation field $H_n$ (*3,4*) of films of Co–Pd alloy as functions of the thermal expansion coefficient of the substrate. All the films had a thickness of 180 Å and Co concentration of 24 at.%. Anneal temperature, °C: *1,3* — 590, *2,4* — 450.

strain (if we measure strain from the state of a free film), which depends only on the film material, elastic stress in the film ceases to be stored and the strain becomes almost completely plastic, and plastic deformation for such values does not change the magnetic anisotropy.

## EXPERIMENTS ON THERMOMAGNETIC RECORDING

The angle of Kerr rotation in samples grown on quartz and optical glass at a wavelength of 630 nm is about 0.2°, which is comparable to the same characteristics for multilayer Co/Pd structures that are analogous with respect to thickness and average composition. In order to reveal the potential of these films as magnetooptic recording media, we carried out experiments on thermomagnetic recording. To write information we used a laser with a wavelength of 1.06 $\mu$m. We used a beam splitter to split the laser beam into two roughly equal-intensity beams and directed each one onto the surface of a magnetic film in order to obtain an interference pattern. Then, after we first magnetized the film to saturation, and illuminated it with laser pulses whose duration was 100 ns and average intensity was $7\times10^4$ W/cm$^2$. In this case the sample was not placed in an external magnetic field. After illumination by laser pulses, the sample was investigated using a polarizing microscope in transillumination. Figure 7 shows the pattern observed in the illuminated region of the film for almost-crossed polarizer and analyzer. The magnetization is in bands that are visible in Fig. 7, and is directed perpendicular to the plane of the figure away from and toward the reader, alternating in direction from the light band to the dark band and conversely. The period of the magnetic structure obtained was 5.7 $\mu$m. It is clear from the fine details that the size of a recording domain can be made considerably smaller than this value. No sign of structural degradation was observed after several cycles of rewriting.

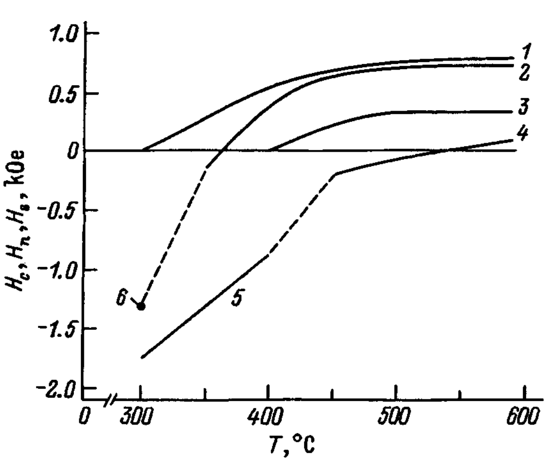

FIG. 6. Dependence of the parameters $H_c$, $H_n$ and $H_s$ on anneal temperature for two films on substrates made of quartz glass and optical glass. The films had a thickness of 180 Å and Co concentration of 25 at.%. *1,2,6* — $H_c$, $H_n$, $H_s$ for a substrate made of quartz glass, *3–5* — $H_c$, $H_n$, $H_s$, respectively, for a substrate made of optical glass.

FIG. 7. Pattern of bands with oppositely directed magnetizations obtained for one of the samples by thermomagnetic recording ($\times$1000).

These results show that under the most favorable conditions (sufficiently small thermal expansion coefficient of the substrate, and a concentration of Co in the film corresponding to the highest value of the sum $K'=(3/2)\lambda\sigma-2\pi M_s^2$ for a given value of elastic stresses) it is possible using the method described in this paper to obtain media with perpendicular magnetic anisotropy that are suitable for magnetooptic recording. The method proposed is distinguished by its simplicity. The magnetic properties of the final structure are insensitive to the method used to deposit the alloy films, due to the last high-temperature anneal. Another advantages of this medium over multilayer structures is its high thermal stability.

Moreover, the results of this work lead us to conclude that a difference in thermal expansion coefficient of the film and substrate can give an additional contribution to the anisotropy connected with elastic stresses. When thermal methods are used to measure the magnetic properties of multilayer Co–Pd structures, the neglect of this contribution can also lead to errors and incorrect interpretations of the measured results.

The authors are deeply grateful to Yu. N. Drozdov for performing the x-ray structural analysis of the films, and also to A. A. Fraerman without whose attention this work would never have been done.

The work was carried out with the financial support of the Russian Fund for Fundamental Research, Grant No. 95-02-05388a.

[1] J. M. MacLaren and R. H. Victora, J. Appl. Phys. **76**, 6069 (1994).

[2] G. A. Bertero, R. Sinclair, C.-H. Park, and Z. X. Shen, J. Appl. Phys. **77**, 3953 (1995).

[3] S. Hashimoto, Y. Ochiai, and K. Aso, J. Appl. Phys. **28**, 1596 (1989).

[4] R. Gontarz, L. Smardz, B. Szymanski, and P. Juzikis, J. Magn. Magn. Mater. **120**, 278 (1993).

[5] D. Weller, H. Brandle, and C. J. Chappert, Magn. Magn. Mat. **121**, 461 (1993).

[6] J. R. Chidress, J. L. Duvail, S. Jasmin *et al.*, J. Appl. Phys. **75**, 6412 (1994).

[7] S. U. Jen and B. L. Chao, J. Appl. Phys. **75**, 5667 (1994).

[8] S. Hashimoto, Y. Ochiai, and K. Aso, J. Appl. Phys. **66**, 4909 (1989).

[9] S. Tsunashima, K. Nagase, K. Nakamura, and S. Uchiyama, IEEE Trans. Magn. **25**, 3761 (1989).

[10] A. D. Akhsakhalyan, Yu. A. Bityurin, S. V. Gaponov *et al.*, Zh. Tekh. Fiz. **52**, 1584 (1982) [ Sov. Phys. Tech. Phys. **27**, 969 (1982)].

[11] I. S. Grigoreva and E. Z. Meĭlikhova (Eds.), *Handbook of Physical Quantities* [in Russian], Energoatomizdat, Moscow, 1991, 1232 pp.

Translated by Frank J. Crowne

# Weakly conductive anisotropic media in an alternating electric field

B. V. Krylov and V. E. Leparskiĭ

*B. I. Stepanov Institute of Physics, Belarus Academy of Sciences, 220602 Minsk, Belarus*

(Submitted August 27, 1996)
Zh. Tekh. Fiz. **68**, 71–74 (April 1998)

The problem of finding the potential function in a weakly conductive anisotropic medium for an assigned distribution of the potential on the boundary and sinusoidal variation with time is solved. It is shown that for each phase of the supply voltage the distribution of the potential is the same as it would be for a static field in an insulating crystal with a nonuniform anisotropy that varies with the phase according to a definite law. © *1998 American Institute of Physics.* [S1063-7842(98)01304-X]

Static and quasistatic electric fields in anisotropic media, including both insulators and conductors, have been thoroughly studied. However, there is an extensive class of materials with low electrical conductivities which behave as conductors with a conductivity $\sigma$ in a low-frequency electric field and as insulators with a dielectric constant $\varepsilon$ in a high-frequency field.[1] Hence it follows that there is an intermediate frequency range, in which both $\varepsilon$ and $\sigma$ influence the distribution of the field in the medium. This regime was experimentally investigated in Ref. 2 in the example of an electrooptic gradient deflector made from a KDP crystal. At room temperature this crystal behaves as a conductor with an anisotropy $\sigma_{\min}/\sigma_{\max}=0.86$ in a static field and as an insulator with $\varepsilon_{\min}/\varepsilon_{\max}=0.5$ at a frequency of 50 Hz. When it is heated to 40–50 °C, it belongs to the intermediate region due to an increase in the 50-Hz conductivity. As a result, a phenomenon is observed wherein the field configuration in the crystal varies as a function of the phase of the applied voltage.

We shall examine the electric fields in crystals specifically for the intermediate frequency range, bearing in mind that a pure insulator $(\omega\to\infty)$ and a pure conductor $(\omega\to 0)$ must be obtained as special cases.

To describe a quasistatic electric field in a weakly conductive medium characterized by the parameters $\varepsilon$ and $\sigma$, we must consider the system of equations

$$\frac{d\rho}{dt}=-\operatorname{div}\,\sigma\,\operatorname{grad}\,\varphi, \tag{1}$$

$$\rho=\operatorname{div}\,\varepsilon\,\operatorname{grad}\,\varphi, \tag{2}$$

where $\varphi$ is the electric field potential and $\rho$ is the bulk charge density.

Homogeneous crystals characterized by $\varepsilon$ and $\sigma$ tensors that do not depend on the coordinates are considered in this paper.

It is also assumed that the boundary conditions for $\varphi$ on the closed boundary $S$ are assigned:

$$\varphi|_s=\varphi(S)\cdot\cos\,\omega t. \tag{3}$$

When $\varepsilon_{ik}$ and $\sigma_{ik}$ are constant, Eqs. (1) and (2) take the forms

$$\frac{\partial\rho}{\partial t}=-\nabla_\sigma^2\varphi,\quad \nabla_\sigma^2=\sigma_{ik}\frac{\partial^2}{\partial X_i\partial X_k}, \tag{4}$$

$$\rho=\nabla_\varepsilon^2\varphi,\quad \nabla_\varepsilon^2=\varepsilon_{ik}\frac{\partial^2}{\partial X_i\partial X_k}. \tag{5}$$

After eliminating the unknown $\rho$ therefrom, we obtain

$$\nabla_\varepsilon^2\frac{\partial\varphi}{\partial t}+\nabla_\sigma^2\varphi=0. \tag{6}$$

Utilizing the fact that (6) is linear and that an harmonic time dependence is assumed in the boundary conditions, we seek a solution in the form

$$\varphi=\phi\,\exp(i\omega t). \tag{7}$$

It is convenient to solve the problem in complex variables, bearing in mind that the physically meaningful potential will be the real or imaginary part of (7). Substituting (7) into (6), we obtain the equation for $\phi$, which is a complex function of the coordinates:

$$(\nabla_\sigma^2+i\omega\nabla_\varepsilon^2)\phi=0. \tag{8}$$

Let us mention some special cases. In a static field $(\omega=0)$ Eq. (8) transforms into $\nabla_\sigma^2\phi=0$, i.e., into the equation for an electrically conductive crystal, as was pointed out in Ref. 1. At large values of $\omega$ or in the case where the conductivity of the crystal is negligibly small $(\sigma=0)$, Eq. (8) gives the usual equation for the electrostatics of an insulating crystal $\nabla_\varepsilon^2\phi=0$. In both cases the equation becomes purely real, and its solution $\phi$ is real-valued.

Below we shall consider the planar problem in the $x,y$ coordinate plane, whose axes correspond to the $X'$ and $Z$ crystallographic axes of the KDP crystal, since we intend to use the approach developed to describe the operation of electrooptic devices (deflectors and lenses) in which the nonuniform fields are usually cylindrical fields, which do not depend on one of the coordinates.

Going over to a coordinate system in which the tensors $\sigma_{ik}$ and $\varepsilon_{ik}$ simultaneously have a diagonal form (this is possible for many crystals), we obtain

$$(\sigma_1+i\omega\varepsilon_1)\,\frac{\partial^2\phi}{\partial x^2}+(\sigma_2+i\omega\varepsilon_2)\,\frac{\partial^2\phi}{\partial y^2}=0. \tag{9}$$

This equation was also considered in Ref. 4.

As we have said, the character of a quasistatic electric field varies with time and depends on the phase of the supply voltage at which it is observed.[2] Let us find the law of this variation. We assume that the solution of Eq. (9) under the assigned boundary conditions (3) has been found and that the physical field of potential (7) has been constructed. Its distribution depends on time. We consider it at the time $t$. We ascertain whether this instantaneous field distribution is similar to any static field distribution that would exist if the region contained a crystal with a suitably chosen anisotropy $\xi$, which is construed as the ratio of the principal values of the dielectric tensor of a nonconducting crystal, or, more precisely, whether there is a real equation

$$\frac{\partial^2\varphi}{\partial x^2}+\xi\frac{\partial^2\varphi}{\partial y^2}=0 \tag{10}$$

which would be satisfied by the field (7). If such an equation exists, what does the effective anisotropy $\xi$ equal at each moment in time?

We substitute (7) into (10) and take the real part

$$\mathrm{Re}\left[\left(\frac{\partial^2\phi}{\partial x^2}+\xi\frac{\partial^2\phi}{\partial y^2}\right)\exp(i\omega t)\right]=0. \tag{11}$$

Since $\Phi$ satisfies Eq. (9), we have

$$\frac{\partial^2\phi}{\partial x^2}=-\frac{\sigma_2+i\omega\varepsilon_2}{\sigma_1+i\omega\varepsilon_2}\,\frac{\partial^2\phi}{\partial y^2}=-\alpha\frac{\partial^2\phi}{\partial y^2}, \tag{12}$$

where

$$\alpha=\frac{\sigma_2+i\omega\varepsilon_2}{\sigma_1+i\omega\varepsilon_1}=\alpha_1+i\alpha_2. \tag{13}$$

Substituting (12) into (11), we obtain

$$\mathrm{Re}\left[(\xi-\alpha)\,\frac{\partial^2\phi}{\partial y^2}\exp(i\omega t)\right]=0. \tag{14}$$

Assuming that $|\partial^2\phi/\partial y^2|\neq 0$ and introducing the notation

$$\vartheta=\arg(\partial^2\phi/\partial y^2), \tag{15}$$

we have

$$\mathrm{Re}[(\xi-\alpha)\exp\, i(\omega t+\vartheta)]=0, \tag{16}$$

whence

$$\xi=\alpha_1-\alpha_2\,\tan(\omega t+\vartheta). \tag{17}$$

We point out some consequences of (17).

1. For each phase of the supply voltage, the field in an assigned region coincides with the electrostatic field which would be induced there if the region were filled by an inhomogeneous medium (in which $\vartheta$ depends on the coordinates) with anisotropy specified by a tensor having the principal values 1 and $\xi$ under the existing boundary conditions.

2. The effective anisotropy $\xi$ varies as a function of the phase of the supply voltage according to (17), i.e., it runs, along with $\tan(\omega t+\vartheta)$, through all values from $-\infty$ to $+\infty$. A similar dependence was discussed in Ref. 4 in the case of a solution of Eq. (9) that was obtained without considering the stationary (in time) boundary conditions.

3. Generally speaking, the effective anisotropy depends on the coordinates, since $\vartheta$ depends on the coordinates.

In the investigation of the electrooptic deflectors described in Ref. 2, an increase in the conductivity of the crystals was achieved by heating. It was found that the ratios of the principal values of the $\varepsilon$ and $\sigma$ tensors varied insignificantly and that only the value of the conductivity changed. Therefore, it would be useful to ascertain how the effective anisotropy $\xi$ varies under the following conditions:

$$\frac{\varepsilon_1}{\varepsilon_2}=\varepsilon=\mathrm{const},\quad \frac{\sigma_1}{\sigma_2}=\sigma=\mathrm{const},\quad \frac{\omega\varepsilon_1}{\sigma_1}=r\in(0,\infty).$$

We represent (13) in the form

$$\alpha_1+i\alpha_2=\frac{\sigma_2-i\omega\varepsilon_2}{\sigma_1-i\omega\varepsilon_1}=\frac{\sigma+\varepsilon}{2}-\frac{\sigma-\varepsilon}{2}\,\frac{1-ir}{1+ir}.$$

Since $|(1-ir)/(1+ir)|=1$, in the range $0\leqslant r<\infty$ the complex number $\alpha_1+i\alpha_2$ runs along a semicircle centered at the point $(\sigma+\varepsilon)/2$ and having a radius $(\sigma-\varepsilon)/2$. The maximum value of $\alpha_2$ is achieved at $r=1$, and the dependence of $\xi$ on the phase of the supply voltage according to Eq. (17) is then most clearly manifested.

Figure 1 presents the variation of $\xi$ as a function of the phase of the supply voltage. The region between curves *2* ($\vartheta=-13.1°$) and *3* ($\vartheta=10.5°$) contains the possible values of $\xi$ at different points of the aperture (the dependence of $\vartheta$ on $x$ and $y$) for $r=1.13$. In addition, curves close to the limiting values $r\to 0$ ($f=0.5$ Hz, curve *1*) and $r\to\infty$ ($f=500$ Hz, curve *4*) are presented to show what happens when one goes over from medium that is described simultaneously by the two tensors $\varepsilon_{ik}$ and $\sigma_{ik}$ to the limiting cases of a pure insulator and a pure conductor.

The pattern of field variations observed in Refs. 2 and 5 in a heated electrooptic deflector can be explained on the basis of the foregoing presentation. Let the deflector electrodes, i.e., conjugate hyperbolas, be designed for a certain anisotropy of the medium. It is seen from Fig. 1 that there is always a phase of the supply voltage in which $\xi$ is close to the anisotropy inherent in the design of the electrodes and that it has a larger or smaller value in other phases. However, all these relations are approximate, since $\xi$ depends on the coordinates (the vertical section of the region between curves *2* and *3*); therefore, the field corresponds only approximately to the field which would induced in the device if a homogeneous electrooptic crystal were placed within the electrode system.

Let us consider the dependence of the effective anisotropy on the coordinates. As follows from (17), $\xi$ depends on the coordinates only if $\vartheta$ depends on them. We ascertain whether there are solutions of Eq. (9) for which $\vartheta=\mathrm{const}$ and, if they exist, what are the boundary conditions for a potential which varies harmonically with time. We note that if this is true for some solution $\Phi$, then $\vartheta=0$ for $\Phi$ exp

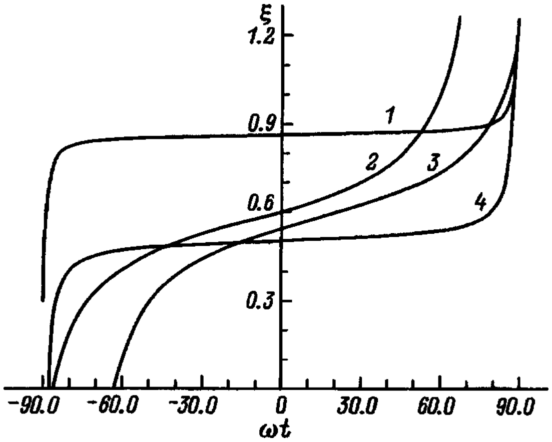

FIG. 1. Dependence of the effective anisotropy $\xi$ on the phase of the supply voltage.

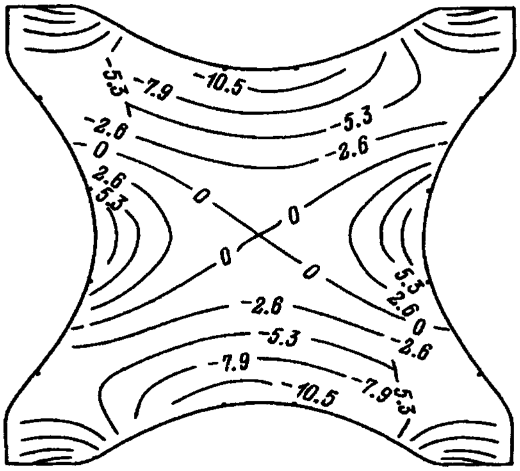

FIG. 2. Lines of equal phase $\vartheta(x,y)$.

$(-i\vartheta)$, which differs by a constant factor and is therefore also a solution of Eq. (9). We seek such solutions.

We take Eq. (9) in form (12)

$$\frac{\partial^2\Phi}{\partial x^2}+\alpha\frac{\partial^2\Phi}{\partial y^2}=0,$$

$$\alpha=\alpha_1+i\alpha_2,\quad \Phi=\Phi_1+i\Phi_2. \tag{18}$$

From

$$\vartheta=\arg\frac{\partial^2\Phi}{\partial y^2}=\arctan\left(\frac{\partial^2\Phi_2}{\partial y^2}\frac{\partial^2\Phi_1}{\partial y^2}\right)=0$$

we obtain

$$\frac{\partial^2\Phi_2}{\partial y^2}=0, \tag{19}$$

whence

$$\Phi_2=F_1(x)y+F_2(x), \tag{20}$$

where $F_1$ and $F_2$ are arbitrary functions at this point.

We separate the real and imaginary parts of Eq. (18). Taking into account (19) and (20), we have

$$\frac{\partial^2\Phi_1}{\partial x^2}=-\alpha_1\frac{\partial^2\Phi_1}{\partial y^2}=\frac{\alpha_1}{\alpha_2}\frac{\partial^2\Phi_2}{\partial x^2}=\frac{\alpha_1}{\alpha_2}[F_1''(x)y+F_2''(x)], \tag{21}$$

i.e.,

$$\Phi_1=\frac{\alpha_1}{\alpha_2}[F_1(x)y+F_2(x)]+f_1(y)x+f_2(y), \tag{22}$$

and, finally, from (21) and (22) we have

$$[F_1''(x)y+F_2''(x)]=-\alpha_2[f_1''(y)x+f_2''(y)]. \tag{23}$$

Since there is a function linear in $y$ on the left-hand side of the equality and a function linear in $x$ on the right-hand side, the functions $F_1''$, $F_2''$, $f_1''$, and $f_2''$ should be linear. After a comparison of the coefficients in front of similar terms and integration, we obtain ($a_i$, $b_j$, $c_k$, and $d_l$ are arbitrary physical constants)

$$\begin{aligned} f_1(y)&=a_1y^3+b_1y^2+c_1y+d_1,\\ f_2(y)&=a_2y^3+b_2y^2+c_2y+d_2,\\ F_1(x)&=-2\alpha_2[a_1x^3+a_2x^2+a_3x+a_4],\\ F_2(x)&=-2\alpha_2[b_1x^3+b_2x^2+b_3x+b_4]. \end{aligned} \tag{24}$$

Classes of solutions of Eq. (18) for which condition (19) is satisfied are demarcated by selecting the arbitrary constants in (24).

In the case of boundary conditions that are symmetric both with respect to $x$ and with respect to $y$ (an electrooptic quadrupole deflector) only even powers of the coordinates can appear in the solution. We then obtain

$$\Phi=b_2(\alpha x^2-y^2)+2\alpha_2b_4-d=A(\alpha x^2-y^2)+B. \tag{25}$$

A similar solution of the equation was considered in Ref. 4, where it was shown that in the case of harmonic variation of the potential with time the shape of the deflector electrodes must vary continually, in contradiction to the boundary conditions of a real device, which are stationary with time. Therefore, it can be stated that condition (19) will not hold for the field existing in a deflector, i.e., $\vartheta$ will depend on the coordinates.

To ascertain the dependence of $\vartheta$ on the coordinates, we must completely solve the problem of finding the electric field potential which satisfies the concrete boundary conditions and Eq. (9). As an example, the field was calculated by the method described in Ref. 6 for a quadrupole deflector with the following initial data: anisotropy inherent in the geometry of the electrodes, 0.79; radius of the circle that is tangent to the electrodes tips, 1.75 mm; temperature, 51.8 °C; control voltage, 2000 V; parameters of the crystal: $\varepsilon_x=42$, $\varepsilon_z=21$, $\varepsilon_0=8.85\times10^{-15}$ F/mm, $\varepsilon_z/\varepsilon_x=0.5$, $\sigma_x=0.6\times10^{-10}$ S/mm, $\sigma_z=0.51\times10^{-10}$ S/mm, $\sigma_z/\sigma_x=0.86$, $\omega\varepsilon_x/\sigma_x=1.13$.

The results of the calculation were used to construct a map (Fig. 2) of lines of equal phase $\vartheta(x,y)$ (the numerical values of the phase of each line are given in degrees) for a phase of the supply voltage equal to zero (the voltage maximum). It is seen from Fig. 2 that the largest phase difference

(~24°) is observed for points in the region near the electrodes. We note that in Fig. 1 the dependence of the anisotropy on the phase is presented for just this case. As follows from Fig. 1, the magnitude of the nonuniformity of the anisotropy (the vertical cross sections of the region between curves *2* and *3*) depends on the phase of the control voltage. The minimum value of the nonuniformity of the anisotropy for devices with the data presented amounts to ~0.06.

This research was financed by the Belarus Foundation for Basic Research.

[1] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 1st ed., Pergamon Press, Oxford (1960) [Russian original, Fizmatgiz, Moscow (1958), p. 303].

[2] N. A. Gusak and A. F. Grib, in *Physical Principles and Practical Aspects of Holography* [in Russian] (1984), pp. 223–229.

[3] B. V. Krylov, Vestsi Akad. Navuk BSSR Ser. Fiz-Mat. Navuk, No. 6, 61 (1978).

[4] N. A. Gusak, Dokl. Akad. Navuk BSSR **28**, 527 (1984).

[5] N. A. Gusak and A. F. Grib, Zh. Tekh. Fiz. **54**, 415 (1984) [Sov. Phys. Tech. Phys. **29**, 249 (1984)].

[6] B. V. Krylov, V. E. Leparskiĭ, and N. I. Gurin, Zh. Tekh. Fiz. **64**(9), 95 (1994) [Tech. Phys. **39**, 912 (1994)].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Modeling of certain nonuniform processes in semiconductor technology

B. A. Zon, S. B. Ledovskiĭ, and A. N. Likholet

*Voronezh State University, 394693 Voronezh, Russia*

(Submitted December 9, 1996)
Zh. Tekh. Fiz. **68**, 75–82 (April 1998)

In this paper two problems are formulated and solved: the problem of diffusion in a two-phase system with a moving boundary, and the problem of taking into account the finite reaction rate for formation of the new phase. A numerical solution is found by a variational method, which has a number of practical advantages. © *1998 American Institute of Physics.* [S1063-7842(98)01404-4]

## INTRODUCTION

The evolution of microelectronics has long required mathematical models of the technological processes used in fabricating semiconductor devices. Although this need for models has been the subject of an extensive literature (see, e.g., Refs. 1–7), there are still a large number of problems that remain unsolved even today. In this paper, two of these problems will be formulated and solved: 1) the diffusion of impurities in a semiconductor under conditions such that formation of a new phase takes place (oxidation, silicide formation, etc.), and 2) formation of a new semiconductor phase by a heterogeneous reaction with a finite rate. Both of these problems are of the Stefanovskiĭ type, although the first of them is linear and the second nonlinear. Both were solved numerically by a variational method, which has a number of advantages in practical problems, notably the fact that the results can be interpreted and described by a relatively small number of plots or tables.

## 1. DIFFUSION OF IMPURITIES DURING A HETEROGENEOUS REACTION

*a) Basic equations.* Let a ‘‘pure’’ semiconductor (phase $a$) occupy the half space $x>0$. The initial distribution of impurities is described by a function $F(x)$. At time $t=0$ the temperature of the sample is raised. For simplicity we will assume that the temperature rise occurs instantaneously. Then a layer of phase $b$ begins to form at the surface of the semiconductor, with a whose thickness that varies according to some function $l(t)$, where $l(0)=0$.

At low temperatures, the diffusion of impurities can be neglected. At high temperatures, impurities begin to diffuse, and it is necessary to determine the distribution of impurities in phases $a$ and $b$ at an arbitrary time. The problem as posed is described by the diffusion equation

$$\frac{\partial C(x,t)}{\partial t}=\frac{\partial}{\partial x}\left(D(x,t)\frac{\partial C(x,t)}{\partial x}\right). \tag{1}$$

Here $C(\cdot)$ is the impurity concentration, which satisfies the initial condition

$$C(x,0)=F(x), \tag{2}$$

and $D(\cdot)$ is the diffusion coefficient of impurities, whose dependence on coordinates and time is connected with the process of formation of phase $b$.

The boundary condition for Eq. (1)

$$\left.\frac{\partial C(x,t)}{\partial x}\right|_{x=0}=0 \tag{3}$$

implies that the surface of the solid is impermeable to impurities. The diffusion coefficient in this model can be written in the form of two equivalent expressions:

$$D(x,t)=D_b+(D_a-D_b)\Theta(x-l(t)), \tag{4}$$

$$D(x,t)=D_a-(D_a-D_b)\Theta(l(t)-x). \tag{5}$$

Here $\Theta(\cdot)$ is the Heaviside unit step function, and $D_a$ and $D_b$, which are constant quantities, are diffusion coefficients for impurities in phases $a$ and $b$. In order to solve this problem we could use the method described in Ref. 8. However, it is simpler to use the method described below. After substituting Eq. (4) into Eq. (1) we obtain

$$\frac{\partial C}{\partial t}-D_b\frac{\partial^2 C}{\partial x^2}=(D_a-D_b)\times\left[\Theta(x-l(t))\frac{\partial^2 C}{\partial x^2}+\delta(x-l(t))\frac{\partial C}{\partial x}\right], \tag{6}$$

where $\delta(\cdot)$ is the Dirac delta function.

Treating the right side of the equation as an inhomogeneous term, let us rewrite this equation in the form of an integrodifferential equation:

$$C(x,t)=\int_0^\infty dy G_b(x,y;t)F(y)+(D_a-D_b)\times\int_0^t dt'\left[G_b(x,l(t');t-t')\left.\frac{\partial C(y,t')}{\partial y}\right|_{y=l(t')}+\int_{l(t')}^\infty dy G_b(x,y;t-t')\frac{\partial^2 C(y,t')}{\partial y^2}\right]. \tag{7}$$

Here $G_b(\cdot)$ is the Green's function for the parabolic operator

$$\frac{\partial}{\partial t}-\frac{\partial}{\partial t}\left(D_b\frac{\partial}{\partial x}\right)$$

with boundary condition (3). The explicit form of this function is well known:[9]

$$G_b(x,y;t)=(4\pi D_b t)^{-1/2}\left\{\exp\left[-\frac{(x-y)^2}{4D_b t}\right]+\exp\left[-\frac{(x+y)^2}{4D_b t}\right]\right\}. \tag{8}$$

In order to transform Eq. (7) we take into account that for $x>l(t)$ the function $C(\cdot)$ satisfies the equation

$$\frac{\partial C}{\partial t}-D_a\frac{\partial^2 C}{\partial x^2}=0,$$

from which it is easy to obtain the relation

$$\left.\frac{\partial C}{\partial x}\right|_{x=l(t)}=-\int_{l(t)}^{\infty}\frac{\partial^2 C}{\partial x^2}dx=-\frac{1}{D_a}\int_{l(t)}^{\infty}\frac{\partial C}{\partial t}dx.$$

Using these relations, let us rewrite Eq. (7) in the following form:

$$C(x,t)=\int_0^\infty dy G_b(x,y;t)F(y)+\frac{D_a-D_b}{D_a}\int_0^t dt'\int_{l(t')}^\infty dy[G_b(x,y;t-t')-G_b(x,l(t);t-t')]\frac{\partial C(y,t')}{\partial t'}. \tag{9}$$

If we use Eq. (5) in the derivation instead of Eq. (4), then Eq. (9) is replaced by

$$C(x,t)=\int_0^\infty dy G_a(x,y;t)F(y)+\frac{D_b-D_a}{D_b}\int_0^t dt'\int_0^{l(t')} dy[G_a(x,y;t-t')-G_a(x,l(t');t-t')]\frac{\partial C(y,t')}{\partial t'}, \tag{10}$$

where $G_a(\cdot)$ differs from $G_b(\cdot)$ by the replacement of $D_b$ by $D_a$.

Equations (9) and (10) can be reduced to integral equations by the substitution

$$P(x,t)=\frac{\partial C(x,t)}{\partial t}.$$

After differentiating Eqs. (9) and (10) with respect to $t$ we obtain

$$P(x,t)=\int_0^\infty dy G_b'(x,y;t)F(y)+\frac{D_a-D_b}{D_a}\int_0^t dt'\int_{l(t')}^\infty dy[G_b'(x,y;t-t')-G_b'(x,l(t');t-t')]P(y,t'), \tag{11}$$

$$P(x,t)=\int_0^\infty dy G_a'(x,y;t)F(y)+\frac{D_b-D_a}{D_b}\times\int_0^t dt'\int_0^{l(t')} dy[G_a'(x,y;t-t')-G_a'(x,l(t');t-t')]P(y,t'). \tag{12}$$

Here

$$G'(\cdot)=\frac{\partial G(\cdot)}{\partial t}.$$

Equations (11) and (12) are the fundamental equations of this section. In deriving them we have not needed to use any matching conditions at the boundary between phases $a$ and $b$. In solving them by the variational method, these conditions will be taken into account in choosing a trial function. Thus these quantitites are of a quite general character. The fact that Eqs. (11) and (12) are mathematically different ways of writing the same equation allows us to monitor the accuracy of our numerical computations by comparing the results obtained for each version of the equation.

Note also that the treatment given here does not include the possibility of a difference in the molar volumes of phases $a$ and $b$. Inclusion of this difference does not change the essence of the method illustrated and it is quite easy to do. This remark also applies to Sec. 2.

*b) Method of numerical calculation.* Equations (11) and (12) were solved numerically by a variational method. In this case, different approximations were used in the different phases, and at the boundary of the phases the following matching conditions were included:

$$C(l(t)-0,t)=C(l(t)+0,t),$$

$$D_b\left.\frac{\partial C}{\partial x}\right|_{x=l(t)-0}=D_a\left.\frac{\partial C}{\partial x}\right|_{x=l(t)+0}. \tag{13}$$

The matching conditions given here are for a completely permeable phase boundary. Assuming that at the initial time the impurities are distributed according to a Gaussian,

$$F(x)=\exp\left(-\left(\frac{x-x_0}{\sigma}\right)^2\right),$$

we approximate $C(\cdot)$ by the following trial function:

$$C(x,t)=N(t)\begin{cases}\exp\left(-\left(\dfrac{x-x_0(t)}{\sigma(t)}\right)^2\right), & x\geqslant l(t),\\ x_0(t)\exp\left(\dfrac{x}{\tilde{\sigma}(t)}\right)^2, & x<l(t).\end{cases} \tag{14}$$

This function depends on five functions of time: $N(t)$, $x_0(t)$, $\sigma(t)$, $\tilde{x}_0(t)$, and $\tilde{\sigma}(t)$, from which $x_0(t)$ and $\sigma(t)$ will be treated as independent and $N(t)$, $\tilde{x}_0(t)$, and $\tilde{\sigma}(t)$ will be determined numerically from the matching condition (13) and the condition of conservation of the number of impurities:

$$\int_0^\infty C(x,t)dx=\int_0^\infty F(x)dx=\text{const}.$$

The form of the trial function $P(\cdot)$ is obtained from the relations

$$P(x,t)=D_a\frac{\partial^2 C(x,t)}{\partial x^2},\quad x\geqslant l(t), \tag{15}$$

$$P(x,t)=D_b\frac{\partial^2 C(x,t)}{\partial x^2},\quad x<l(t), \tag{16}$$

which follow directly from Eq. (1) and Eqs. (4) or (5). Substituting Eq. (14) into Eq. (15) and (16), we find the following form of $P(\cdot)$:

$$P(x,t)=N(t)\begin{cases}\dfrac{2D_a}{\sigma^2(t)}\exp\left(-\left(\dfrac{x-x_0(t)}{\sigma(t)}\right)^2\right)\\ \qquad\times\left[\dfrac{2(x-x_0(t))^2}{\sigma^2(t)}-1\right], & x\geqslant l(t),\\ \dfrac{2D_b}{\tilde{\sigma}^2(t)}x_0(t)\exp\left(\dfrac{x}{\tilde{\sigma}(t)}\right)^2\\ \qquad\times\left[\dfrac{2x^2}{\tilde{\sigma}^2(t)}+1\right], & x<l(t).\end{cases}$$

The calculation was optimized by the method of the golden section from the condition of minimization of the norm of Eq. (11),

$$R_a(t)=\int_0^\infty Q_a^2(x,t)dx,$$

where

$$Q_a(x,t)=\int_0^\infty dyG_b'(x,y;t)F(y)+\frac{D_a-D_b}{D_a}\times\int_0^t dt'\int_{l(t')}^\infty[G_b'(x,y;t-t')-G_b'(x,l(t');t-t')]P(y,t')dy-P(x,t).$$

In order to monitor the accuracy of the calculations we also minimized the norm of Eq. (12),

$$R_b(t)=\int_0^\infty Q_b^2(x,t)dx,$$

where

$$Q_b(x,t)=\int_0^\infty dyG_a'(x,y;t)F(y)+\frac{D_b-D_a}{D_b}\int_0^t dt'\int_0^{l(t')}[G_a'(x,y;t-t')-G_a'(x,l(t');t-t')]P(y,t')dy-P(x,t).$$

The results of the calculation, which depend on the relation between the diffusion coefficients of phases $a$ and $b$, are given in the next section.

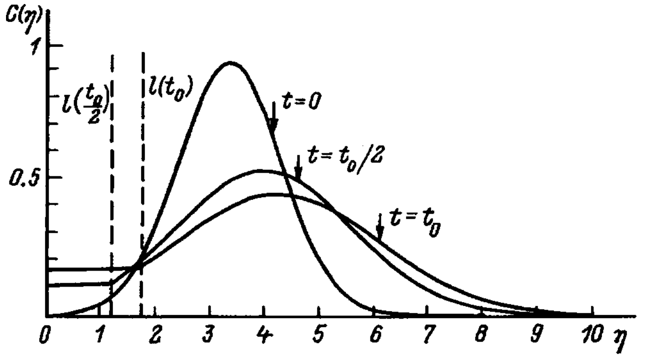

FIG. 1. Depth profile of impurity concentration at various times measured from the start of silicide formation for $D_a\ll D_b$.

*c) Numerical results and discussion.*

1) $D_a\ll D_b$. A typical example of this case is provided by diffusion of phosphorous during the formation of titanium silicide TiSi. For $T=600$ °C the diffusion coefficients of phosphorous are $D_a=D_{\mathrm{Si}}=0.26\times10^{-20}$ cm$^2$/s, $D_b=D_{\mathrm{TiSi}}=8\times10^{-17}$ cm$^2$/s .[1,10] The results of the calculations are shown in Fig. 1, where $t_0$ is the time it takes the impurity concentration to decrease from its maximum by a factor of 2, and $\eta=x/\sqrt{4\pi D_a t_0}$ is a dimensionless variable having the meaning of a length. The vertical dotted lines indicate the position of the boundary between phases. In the calculations it was assumed that the time dependence of the boundary is a parabolic function. In units of $\eta$,

$$l(t)=\sqrt{\frac{\pi t}{t_0}}.$$

The smooth profile for the distribution and accumulation of impurities in phase $b$ is the result of the large ratio of diffusion coefficients $D_b/D_a=3\times10^4$. Similar behavior of impurities is observed experimentally in Refs. 11–13.

2) $D_a\gg D_b$. This situation arises during the formation of silicides of the platinum metal group (Pd and Pt) for a wide class of impurities (P, As, Sb, etc.).[11,13–15] As a computational example the following values of diffusion coefficients are considered: $D_a=0.26\times10^{-20}$ cm$^2$/s, $D_b=0.8\times10^{-24}$ cm$^2$/s. The results of the calculations are shown in Fig. 2. It is clear that this case corresponds to outflow of impurities from phase $b$.

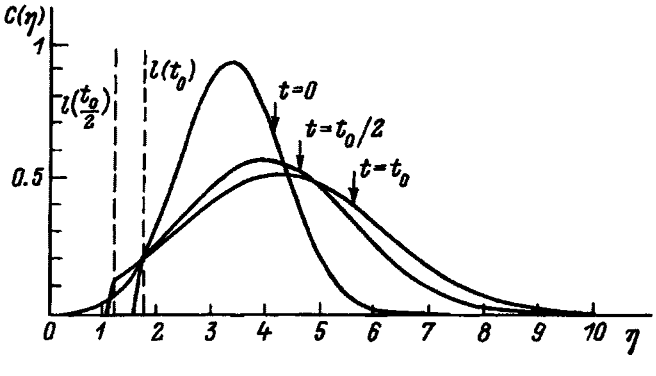

FIG. 2. The same as in Fig. 1 for $D_a\gg D_b$.

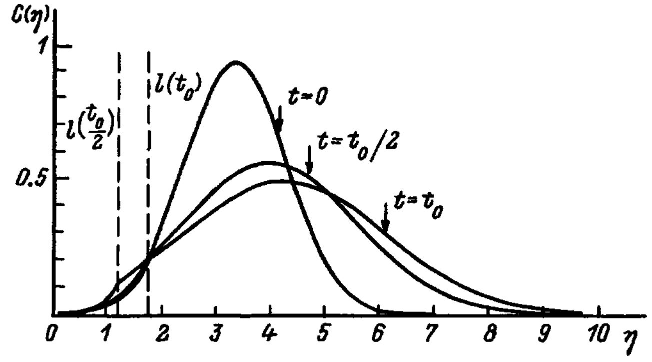

FIG. 3. The same as in Fig. 1 for $D_a \sim D_b$.

3) $D_a \sim D_b$. This case is rarely encountered in the theory of impurity diffusion in semiconductors; however, it is treated here because the results obtained could be useful in solving other problems, for example in the theory of thermal conductivity. The results of these calculations are shown in Fig. 3. The following case is considered: $D_a = 0.26 \times 10^{-20}$ cm$^2$/s, $D_b = 0.5 \times 10^{-20}$ cm$^2$/s.

In the special case where the diffusion constants are not very different ($|D_a - D_b| \ll D_a, D_b$), a small parameter appears in the theory. In this case Eqs. (9) and (10) can be solved by the method of successive approximations. We will not pause to discuss this question.

## 2. EFFECT OF FINITE RATE OF CHEMICAL REACTION ON THE STRUCTURE OF THE PHASE BOUNDARY

*a) Mathematical simulation of the process.* For definiteness let us discuss a reaction that forms a silicide; we will assume that the silicide forms as a result of diffusion of the metal into silicon. Let us assume that at the beginning of the process the metal occupies the space $x < 0$, and silicon occupies the space $x > 0$. Let us write the equation that describes the process in the following form:

$$\frac{\partial M}{\partial t} = \frac{\partial}{\partial x}\left( D_M \frac{\partial M}{\partial x} \right) + S, \tag{17}$$

where $M$ is the concentration of free metal atoms, $D_M$ is the diffusion coefficient fo the metal, which in general depends on coordinate and time, and $S$ is the density of the ''sink'' of metal atoms associated with the possibility of silicide formation.

The boundary and initial conditions for Eq. (17) have the general form

$$M(0,t) = n_0, \quad M(x,0) = 0, \quad x > 0, \tag{18}$$

where $n_0$ is the density of atoms in the pure metal.

In order to calculate the quantity $S$ let us consider the rate of formation of the silicide. Let $n_a$ be the concentration of free silicon atoms, $n_b$ the silicide concentration. Then

$$\frac{\partial n_b}{\partial t} = k n_a M, \tag{19}$$

where $k$ is the rate constant of the reaction.

Here we will treat the simplest case, where the stable silicide has the chemical formula MeSi. Note that the well-known parabolic regime of kinetics[5] arises in the limit of an infinitely rapid rate of the chemical reaction $k \to \infty$. The goal of this section is to take into account the finiteness of the value of $k$. Neglecting the possibility of the release or absorption of heat during the reaction, we assume that the reaction rate constant $k$, like the diffusion coefficient $D_M$, does not depend explicitly on time. It is easy to see that

$$S = \left( \frac{\partial M}{\partial t} \right)_R = -\frac{\partial n_b}{\partial t}, \tag{20}$$

where $(\partial M / \partial t)_R$ denotes the change in the number of metal atoms in a certain small region of space due to the reaction.

Assuming that there is no change in the volume of the solid during the reaction, along with Eq. (19) we can write the following relation

$$n_a + n_b = n_{a0}, \tag{21}$$

where $n_{a0}$ is the concentration of silicon atoms in the ''pure'' silicon.

Equation (21) expresses the law of conservation of silicon atom density, which holds in the absence of silicon diffusion. Substituting Eq. (21) into Eq. (19), we obtain

$$\frac{\partial n_b}{\partial t} = k(n_{a0} - n_b) M. \tag{22}$$

The solution to this equation with the condition $n_b(x,0) = 0$ has the form

$$n_b(x,t) = n_{a0} \left\{ 1 - \exp\left[ -k \int_0^t M(x,t')\, dt' \right] \right\}. \tag{23}$$

Substituting this expression into Eqs. (20) and (17), we find an equation for the distribution of free metal in the region $x > 0$:

$$\frac{\partial M(x,t)}{\partial t} = \frac{\partial}{\partial x}\left( D_M \frac{\partial M(x,t)}{\partial x} \right) - k n_{a0} M(x,t) \times \left\{ \exp\left[ -k \int_0^t M(x,t')\, dt' \right] \right\}. \tag{24}$$

Equation (24) is a nonlinear integrodifferential equation; therefore, its analytical investigation is quite difficult. A significant simplification of the equation is obtained if we assume that the diffusion coefficients of the metal in pure silicon and in the silicide are the same. It is this case that we now discuss.

*b) Analytical treatment of the case $D_M$=const.* We introduce the dimensionless variables

$$\rho^2 = \frac{n_{a0}}{n_0}, \quad \tau = k n_0 t, \quad \xi = \left( \frac{k n_0}{D_M} \right)^{\frac{1}{2}} x,$$

$$g(\xi,\tau) = \frac{M(\xi,\tau)}{n_0}, \tag{25}$$

and rewrite Eq. (24) and the subsidiary condition (18) as

$$\frac{\partial g(\xi,\tau)}{\partial \tau} = \frac{\partial^2 g(\xi,\tau)}{\partial \xi^2} - \rho^2 g(\xi,\tau) \exp\left[ -\int_0^\tau g(\xi,\tau')\, d\tau' \right], \tag{26}$$

$$g(0,\tau)=1, \quad g(\xi,0)=0.$$

We can transform the integrodifferential equation (26) into a differential equation by introducing a new dependent variable

$$f(\xi,\tau)=\int_0^{\tau} g(\xi,\tau')\,d\tau'.$$

The equation for $f(\xi,\tau)$ has the form of the Kolmogorov–Petrovskiĭ–Piskunov equation

$$\frac{\partial f}{\partial \tau}=\frac{\partial^2 f}{\partial \xi^2}+\rho^2(\exp(-f)-1), \tag{27}$$

$$f(0,\tau)=\tau, \quad f(\xi,0)=0. \tag{28}$$

In this case the additional condition

$$g(\xi,0)=\frac{\partial f}{\partial \tau}\Big|_{\tau\to 0}$$

is satisfied automatically. An analytical solution to the boundary value problem (27) and (28) can be found in two limiting cases.

1) The region $f\leqslant 1$. The conditions for physical realization of this region will be clarified below. Expanding $\exp(-f)$ in Eq. (27) in a series and retaining the first two terms, we obtain a linear equation

$$\frac{\partial f}{\partial \tau}=\frac{\partial^2 f}{\partial \xi^2}-\rho^2 f. \tag{29}$$

The solution to this equation with the subsidiary condition (28) has the following form:

$$f(\xi,\tau)=\frac{1}{2}\left\{\left(\tau-\frac{\xi}{2\rho}\right)\exp(-\rho\xi)\left[1+\Phi\left(\rho\sqrt{\tau}-\frac{\xi}{2\sqrt{\tau}}\right)\right]\right.$$
$$\left.+\left(\tau+\frac{\xi}{2\rho}\right)\exp(\rho\xi)\left[1-\Phi\left(\rho\sqrt{\tau}+\frac{\xi}{2\sqrt{\tau}}\right)\right]\right\}. \tag{30}$$

Here $\Phi(\cdot)$ is the probability integral.[9] Equation (30) simplifies when

$$\xi\ll\sqrt{\tau}\ll 1. \tag{31}$$

In this case

$$f(\xi,\tau)\simeq\left(1-\frac{\xi}{\sqrt{\pi\tau}}\right)\tau. \tag{32}$$

It is not difficult to see that condition (31) applies during the initial stage of the reaction near the metal–silicon boundary. In this region the density of metal atoms changes according to a linear law:

$$g(\xi,\tau)\simeq 1-\frac{\xi}{\sqrt{\pi\tau}}. \tag{33}$$

As is well known, the probability integral very rapidly reaches its asymptotic value. Therefore, when the condition

$$\xi\gg\sqrt{\tau}, \quad \xi>2\rho\tau \tag{34}$$

holds, expression (30) also simplifies:

$$f(\xi,\tau)\simeq\frac{8\tau^{\frac{5}{2}}}{\sqrt{\pi}\xi^3}\left(1-\frac{4\rho^2\tau^2}{\xi^2}\right)^{-2}\exp\left(-\rho^2\tau-\frac{\xi^2}{4\tau}\right),$$

$$g(\xi,\tau)\simeq\frac{2\sqrt{\tau}}{\sqrt{\pi}\xi}\left(1-\frac{4\rho^2\tau^2}{\xi^2}\right)^{-1}\exp\left(-\rho^2\tau^2-\frac{\xi^2}{4\tau}\right). \tag{35}$$

Condition (34), and consequently Eq. (35), are valid far from the phase boundary beween silicide and silicon, where the metal atom density is small. Finally, in the region

$$\xi\gg\frac{1}{\rho}, \quad \xi\gg\sqrt{\tau}, \quad \xi<2\rho\tau \tag{36}$$

we have

$$f(\xi,\tau)\simeq\left(\tau-\frac{\xi}{2\rho}\right)\exp(-\rho\xi), \quad g(\xi,\tau)\simeq\exp(-\rho\xi). \tag{37}$$

The region (36) is found to be close to the silicide–silicon boundary; therefore the density of metal atoms falls off here according to an exponential law, whereas in region (34) it falls off like a Gaussian.

Expressions (35) and (37) become meaningless for $\xi=2\rho\tau$. In this case,

$$f\left(\xi,\tau=\frac{\xi}{2\rho}\right)\simeq\begin{cases}\dfrac{\xi}{2\rho}-\left(\dfrac{2\xi}{\pi\rho}\right)^{\frac{1}{2}}, & \xi\ll 1,\\[2ex] \left(\dfrac{\xi}{\rho}(2\rho)^3\right)^{-\frac{1}{2}}\exp(-\rho\xi), & \xi\gg 1,\end{cases}$$

$$g\left(\xi,\tau=\frac{\xi}{2\rho}\right)\simeq\begin{cases}\dfrac{1}{2}-\left(\dfrac{2\rho\xi}{\pi}\right)^{\frac{1}{2}}, & \xi\ll 1,\\[2ex] \left(\dfrac{1}{2}+(8\pi\rho\xi)^{-\frac{1}{2}}\right)\\ \quad\times\exp(-\rho\xi), & \xi\gg 1.\end{cases} \tag{38}$$

2) The region $f\gg 1$. Here we can neglect the exponential in Eq. (27) and write the solution to the boundary value problem in the following form:

$$f(\xi,\tau)\simeq\left(\tau+\frac{\xi^2}{2}\right)\left[1-\Phi\left(\frac{\xi}{2\sqrt{\tau}}\right)\right]-\frac{\xi\sqrt{\tau}}{\sqrt{\pi}}\exp\left(-\frac{\xi^2}{4\tau}\right)$$
$$-\rho^2\int_0^{\tau}d\tau'\,\Phi\left(\frac{\xi}{2\sqrt{\tau-\tau'}}\right). \tag{39}$$

In the steady-state regime near the metal–silicide boundary for

$$\tau\gg 1, \quad \xi^2\ll\tau \tag{40}$$

we can obtain the following expressions:

$$f(\xi,\tau)\simeq\tau\left(1-\frac{2\xi}{\sqrt{\pi\tau}}\right), \quad g(\xi,\tau)\simeq 1-\frac{\xi}{\sqrt{\pi\tau}}. \tag{41}$$

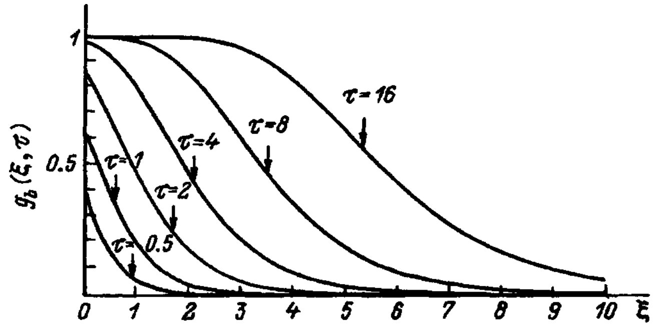

FIG. 4. Distribution of silicide concentration with respect to depth at various times measured from the beginning of the process.

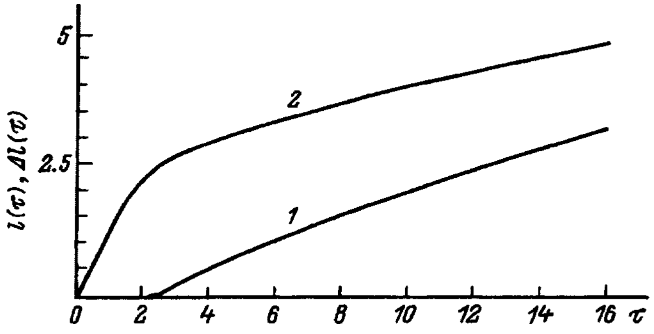

FIG. 5. Time dependence of the silicide thickness (*1*) and transition layer thickness (*2*).

In obtaining the latter expressions, we estimated the last term in Eq. (39) in the following way. After the change of variable $\xi/(2\sqrt{\tau-\tau'})=s$, the function $\Phi(\cdot)$ is expanded in a series:

$$\Phi(s)=\frac{2}{\sqrt{\pi}}\exp(-s^2)\sum_{i=0}^{\infty}\frac{2^i s^{2i+1}}{(2i+1)!!} \tag{42}$$

and is integrated term by term. As in case (33), the metal-atom density in this limit follows a linear law. In order to find the dimensionless concentration of silicide $g_b=n_b/n_0$ in each of the cases discussed here, it is sufficient to substitute $f(\cdot)$ from Eqs. (30), (32), (35), (37)–(39), or (41) into the expression

$$g_b(\xi,\tau)=1-\exp(-f(\xi,\tau)). \tag{43}$$

Thus, our investigation of the linearized equation (27) allows us to find $g(\xi,\tau)$ and $g_b(\xi,\tau)$ near the metal–silicide boundary and far from the silicide–silicon boundary. In order to find the distribution of metal and silicide concentrations near the silicide–silicon boundary, it is necessary to resort to numerical methods even for $D_M=$ const.

*c) Method of numerical calculations.* For $f(\cdot)$ we chose the expression

$$f(\xi,\tau)=\tau\exp(-\alpha(\tau)\xi^{\beta(\tau)}) \tag{44}$$

which, starting from the boundary and initial conditions and the analytical approximations, was used for numerical solution of Eq. (27).

The numerical dependences of $\alpha(\tau)$ and $\beta(\tau)$ were found by the method of optimization. For this, the entire range of times of interest to us from the beginning of the process was divided up into small intervals $|\tau_i,\tau_{i+1}|$, within each of which $\alpha$ and $\beta$ may be treated as constants. The quantity

$$\mathrm{dif}(\xi,\tau_i,\tau_{i+1})\equiv\int_{\tau_i}^{\tau_{i+1}}\frac{\partial^2 f}{\partial\xi^2}d\tau-f(\tau_{i+1})+f(\tau_i)+\rho^2 \times\left(\int_{\tau_i}^{\tau_{i+1}}\exp(-f(\tau))d\tau-(\tau_i-\tau_{i+1})\right)$$

after substitution of $f(\xi,\tau)$ from Eq. (44) and replacing the integral by its approximate value based on the trapezoid formula, can be written in the form

$$\begin{aligned}\mathrm{dif}(\xi,\tau_i,\tau_{i+1})\cong{}&\frac{\tau_{i+1}-\tau_i}{2}\{\exp(-\alpha\xi^\beta)[\alpha\beta\xi^{\beta-2}\\&\times(\alpha\beta\xi-\beta-1)(\tau_i+\tau_{i+1})-2]\\&-\rho^2[\exp(-\tau_{i+1}\exp(-\alpha\xi^\beta))\\&+\exp(-\tau_i\exp(-\alpha\xi^\beta))-2]\}.\end{aligned}$$

On each interval $[\tau_i,\tau_{i+1}]$ the values of $\alpha((\tau_i+\tau_{i+1})/2)$ and $\beta((\tau_i+\tau_{i+1})/2)$ were determined by the method of the ''golden section'' by minimizing the quantity

$$S_{i,i+1}\equiv\int_0^\infty \mathrm{dif}^{\,2}(\xi,\tau_i,\tau_{i+1})d\xi. \tag{45}$$

Once $\alpha$ and $\beta$ were found, the values of $f$ and $g_b$ were computed from Eqs. (44) and (43) and the concentration of metal was computed from the expression

$$g(\xi,\tau)=\frac{\partial f}{\partial\tau}. \tag{46}$$

*d) Results of calculations and discussion.* In this section we present the results of numerical solution to Eq. (27) using the method described above, for $\tau_i$ ranging from 0 to 16 and for $\rho=1$. From these calculations it follows that $\beta\approx1$ under these conditions, which allows us in principle to limit consideration to just one function to be optimized, $\alpha(\tau)$. However, it was not possible to confirm this ahead of time; therefore, for generality in the calculations we optimized both $\alpha(\tau)$ and $\beta(\tau)$. Figure 4 shows the silicide concentration profiles at different times measured from the beginning of the reaction. It is clear that for small times the silicide concentration is smaller than the concentration of atoms of the original material even at the surface, and it falls off monotonically with depth in the crystal. This is a consequence of including the finite rate of reaction. As the reaction goes on, a silicide layer is formed in the subsurface region, with a constant concentration equal to the concentration of the original material. Behind it follows a transition layer of variable composition and then the original material. At still later times in the process both the thickness of the silicide layer and the thickness of the transition layer increase (we take the thickness of the silicide layer to be the distance from the surface to a plane with a silicide concentration of 0.9 of the maximum and the thickness of the transition layer to be the distance between planes with concentrations of 0.9 and 0.1 of the maximum). In this case, the time dependence of the silicide layer thickness $l$ (Fig. 5, curve *1*) is closer to linear than to $l \sim \sqrt{\tau}$, the dependence that is characteristic of a model of the process that does not take the finite reaction rate for silicide formation into account. The time dependence of the transition layer thickness $\Delta l$ (Fig. 5, curve *2*) consists of two segments, both of which are close to linear. The first segment of more rapid growth corresponds to the stage in which the concentration of silicide at the surface has not yet saturated. This stage is characterized by the induction time of the reaction. After the surface concentration saturates, a layer of silicide begins to form on the surface side of the transition region, resulting in a significant slowing of the increase in the transition layer thickness. To summarize, the model described in this paper not only leads to a more correct determination of the rate of silicide formation but also allows one to take into account the presence of a transition layer and to compute its parameters as functions of the reaction conditions.

[1] *Fundamentals of Silicon Integrated Device Technology. Vol. 1: Oxidation, Diffusion and Epitaxy*, R. M. Burger and R. P. Donovan (Eds.) [Prentice Hall, New Jersey, (1967); Mir, Moscow (1969)].

[2] J. M. Poate, K. N. Tu, and J. M. Mayer (Eds.), *Thin Films — Interdiffusion and Reactions*, [New Jersey (1978); Mir, Moscow (1982)].

[3] S. P. Murarka, *Silicides for VLSI Applications* [Academic Press, New York (1983); Mir, Moscow (1986)].

[4] V. P. Maslov, V. G. Danilov, and K. A. Volosov, *Mathematical Modeling of Thermal Mass-Transfer Processes* [in Russian], Nauka, Moscow, 1987.

[5] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits* [Wiley, New York (1986); Mir, Moscow (1989)].

[6] A. E. Gershinskiĭ, A. V. Rzhanov, and B. I. Cherepov, Poverkhnost', No. 2, 1 (1982).

[7] S. V. Vasil'ev and N. N. Gerasimenko, Poverkhnost', No. 7, 57 (1986).

[8] B. Ya. Lobov, *Theory of Crystallization in Large Volumes* [in Russian], Nauka, Moscow, 1975.

[9] A. N. Tikhonov and A. A. Samarskiĭ, *Equations of Mathematical Physics* [in Russian], 4th ed. (Nauka, Moscow, 1972).

[10] P. Gas, G. Scilla, A. Michel *et al.*, Appl. Phys. **63**, 5335 (1988).

[11] M. Wittmer, C. Y. Ting, and K. N. Tu, Thin Solid Films **104**, 215 (1983).

[12] C. J. Wei, W. Katz, and C. Smitz, Thin Solid Films **104**, 215 (1983).

[13] M. Wittmer and T. E. Seidel, Appl. Phys. **49**, 5827 (1988).

[14] I. Ohdomari, K. N. Tu, K. Sugaro *et al.*, Appl. Phys. Lett. **38**, 1015 (1981).

[15] P. L. Thornton, Electron. Lett. **17**, 480 (1981).

Translated by Frank J. Crowne

# Distinctive features of time-resolved photoluminescence of porous silicon coated with a layer of diamondlike carbon

Yu. P. Piryatinskiĭ

*Institute of Physics, National Academy of Sciences of Ukraine, 252028 Kiev, Ukraine*

V. A. Semenovich

*Institute of Ultrahard Materials, National Academy of Sciences of Ukraine, 254074 Kiev, Ukraine*

N. I. Klyuĭ and A. G. Rozhin

*Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, 252028 Kiev, Ukraine*

(Submitted November 29, 1996)
Zh. Tekh. Fiz. **68**, 83–87 (April 1998)

Time-resolved photoluminescence from porous silicon coated with a diamondlike carbon film is investigated. The intensity of the photoluminescence from the carbon film is obserd to increase after deposition, and there is an accompanying change in the intensity and a short-wavelength shift of the photoluminescence band of porous silicon that depends on the porosity of its original layers. These changes are explained by the formation of carbon nanoclusters on the surface of the silicon filaments. © *1998 American Institute of Physics.* [S1063-7842(98)01504-9]

## INTRODUCTION

The prospect of creating light-emitting elements based on porous silicon for silicon-based optoelectronics[1,2] has attracted considerable research interest in studying the properties of this material. The ease of obtaining porous silicon and its intense photoluminescence (PL) in the visible region of the spectrum has led to the publication of a large number of papers on this subject.[1–4] In the course of this activity, a number of hypotheses have been put forward in order to explain the PL of porous silicon, involving assumptions about size quantization of charge carriers localized in thin silicon filaments,[1,2] chemabsorption of molecules at the porous silicon surface and the formation of siloxane, silicon hydrides, etc.,[5,6] and finally radiation at the silicon–adsorbate boundary.[7] Unfortunately, the low degradation resistance of porous silicon[5] significantly restricts its potential usefulness in applications. For this reason, there is much current interest in work directed, on the one hand, towards finding ways to increase the ruggedness of porous silicon and, on the other hand, to provide additional information that would aid in understanding the mechanism responsible for the visible PL of porous silicon. As was mentioned in Ref. 8, one way to obtain such information is the controlled removal (or deposition) of an adsorbate from (or onto) the surface of the silicon filaments. Such an adsorbate could be a gas, such as molecular chlorine,[8] carbon in the form of fullerene molecules implanted in the porous silicon,[9] or some other material.

In light of the results reported in Refs. 8 and 9, there is special interest in studying the effect of carbon on the photoluminescence of porous silicon. Thus, in Ref. 8 it was established that increasing the carbon content of porous silicon can lead under annealing to agglomeration of the carbon with the formation of graphite nanoclusters, which significantly modify the PL spectrum of porous silicon. On the other hand, it was shown in Ref. 9 that under ordinary conditions the weak luminescence of the $C_{60}$ molecule is strongly enhanced when it is embedded in porous silicon, due to transfer of charge carriers from the silicon nanocrystallites to the $C_{60}$ molecules adsorbed on their surfaces. At the same time, the effect of carbon in other forms, for instance in its diamondlike modification, on the properties of porous silicon remain unstudied. It should also be noted that in the majority of studies of the luminescence properties of porous silicon[1–9] traditional methods are used to measure the stationary spectrum of the photoluminescence. It is obvious that by using time-resolved methods of measurement of the PL spectra,[10] especially in studying the porous silicon-adsorbate systems, one should be able in a number of cases to discriminate between contributions to the resultant PL spectra from various sources of radiation, and consequently obtain additional information about the PL mechanisms for these systems.

The goal of this paper is to investigate the distinctive features of time-resolved PL from porous silicon coated with a layer of diamondlike carbon.

## EXPERIMENTAL METHOD

Samples of porous silicon were obtained by the standard method of electrochemical etching of silicon. Wafers of $p$-Si with a resistivity of $10\Omega\cdot$cm (KDB-10) and with the (100) orientation were used. Prior to preparation of the porous silicon, aluminum Ohmic contacts were deposited on the back of these wafers, and then a layer of porous silicon was formed in a Teflon cell with a platinum electrode. The etching took place in darkness in a solution of 48%HF : $C_2H_5OH$

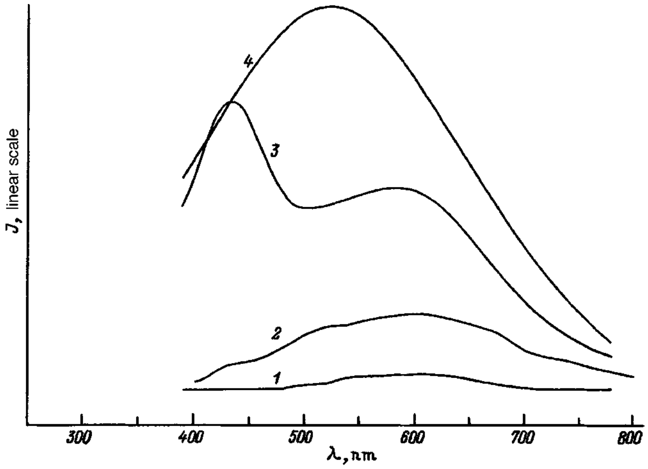

FIG. 1. Photoluminescence spectrum of diamondlike film, porous silicon, and porous silicon plus diamondlike film with nanosecond resolution. In Fig. 1, just as in Figs. 2, 3, and 4 the scale along the ordinate is linear, and *l* is the wavelength.

with in a 1:1 ratio. The samples of porous silicon were prepared using current densities of 10–75 mA/cm$^2$ and etch times of 1.5–10 minutes.

The films of diamondlike carbon were deposited using a capacitive rf discharge plasma (13.56 MHz) at low pressures (0.8 Torr) and a room-temperature substrate (300 K). A gaseous mixture of $CH_4 : H_2 : N_2$ was used. During the deposition the substrate was subjected to an rf potential equal to 1900 V.[11] The thickness of the diamondlike film was measured using an LÉF-3M laser ellipsometer at a wavelength of 632.8 nm, and ranged from 50 to 100 nm.

Time-resolved PL spectra of porous silicon, the diamondlike film, and the combined system were measured using the method described in Ref. 10. A nitrogen laser ($\lambda =$ 337.1 nm) was used for excitation, with an excitation pulse duration of 10 ns and an in-pulse power of 3 kW. A stroboscopic system was used for recording, allowing us to measure the PL spectrum and investigate the kinetics of its changes. PL spectra were recorded with nanosecond and microsecond time delays ($t_d$) relative to the maximum of the laser pulse at room temperature.

## RESULTS AND DISCUSSION

Figure 1 shows the nanosecond PL spectrum of a diamondlike carbon film deposited on a quartz substrate (curve *2*, $d = 70$ nm) and on an unetched Si surface (curve *1*, $d = 70$ nm). Here we also plot the nanosecond PL spectra of porous silicon (curve *3*, $j = 75\,\mathrm{mA/cm^2}$, $t = 5$ min) and the combined system ($d = 70$ nm, curve *4*). It is seen from Fig. 1 (curve *2*) that the PL spectrum of the diamondlike film deposited on a quartz substrate has a band with a maximum at 600 nm and with shoulders on its short-wavelength side at 500 and 440 nm. However, the intensity $J$ of PL from a thin diamondlike layer ($d = 70$ nm) deposited on the unetched surface of Si (Fig. 1, curve *1*) is considerably smaller than for the diamondlike film deposited on a quartz substrate, and its spectrum has no structure. These spectral characteristics of diamondlike films on the surface of silicon could be due to the presence of nonradiative recombination centers at the boundary between the silicon and the diamondlike film. It should also be noted that the PL of diamondlike films is short-lived and appears only in spectra measured with nanosecond resolution. One of the reasons for this could be the formation of tightly bound electron–hole pairs of small radius, which occurs on account of the low dielectric constant of the diamondlike carbon film.[12,13] In measuring the PL spectrum of porous silicon with nanosecond resolution we observe weak emission at wavelengths shorter than 590 nm. During the PL measurements when $t_d = 0$, i.e., at the maximum of the laser pulse, the PL maximum is located around 440 nm (Fig. 1, curve *3*). The intensity of this PL maximum depends on the conditions under which the porous silicon was obtained, and is usually observed after long etching. Moreover, a band at 600 nm also appears in the spectrum, whereas in integrated PL spectra for porous silicon emission in the range $\lambda >$ 600 nm usually predominates.[8,9]

Deposition of a thin ($d = 70$ ns) diamondlike film on the porous silicon surface leads to a shift in the band from 600 to 540 nm and a sharp increase in its intensity (Fig. 1, curve *4*). In this case the band at 440 nm observed in the PL spectrum of porous silicon disappears.

In our opinion, the observed changes are due to penetration of carbon into the pores of the porous silicon during the deposition of the diamondlike film and formation of carbon nanoclusters at the surface of the silicon filaments.[8] The considerable increase in intensity of the PL band that is characteristic of diamondlike films (at 600 nm) is probably due to transfer of charge carriers from the silicon filaments to the carbon clusters adsorbed on their surfaces. In this case the porous silicon plays the role of a generator of charge carriers, and the carbon clusters act as centers of radiative recombination analogous to surface states in the model proposed by Koch and co-authors[7] and used by the authors of Ref. 9 to explain the anomalous increase in PL intensity from fullerene molecules adsorbed in porous silicon layers. The short-wavelength shift we observe in the PL bands from 600 to 540 nm is probably a manifestation of quantum size effects, and its value can depend on the dimensions of the carbon clusters.

A somewhat different character appears when the PL spectrum of porous silicon coated with a diamondlike film of thickness $d =$ 70 nm is measured with microsecond time delays (Fig. 2). We emphasize that for the reasons described above, the diamondlike carbon films do not luminesce in the microsecond range, i.e., the spectra shown in Fig. 2 are caused by luminescence of the porous silicon layer itself and the material characterized as porous silicon plus diamondlike film. It is clear from Fig. 2 that an intense band is present in the microsecond spectra for PL of porous silicon at 680 nm. As the delay time is increased from 0 to 30 $\mu$s, the maximum of this band shifts towards the long-wavelength region of the spectrum out to 720 nm (Fig. 2, curves *1–6*). This could be related to a scatter in the size of the silicon filaments.[1] On this figure we show PL spectra of porous silicon after deposition of a diamondlike carbon film (Fig. 2; curves *1′–6′*) were measured with the same time delays 0–30 $\mu$s. It is clear that after deposition of the diamondlike film the spec-

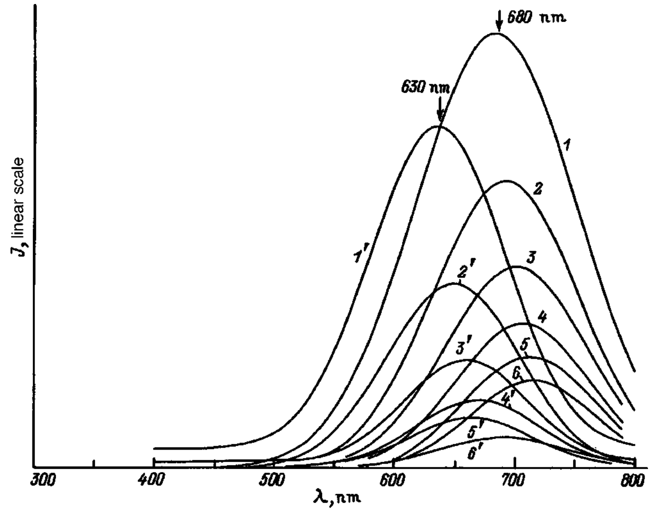

FIG. 2. Photoluminescence spectra (*1–6*) of porous silicon plus diamondlike film (*1′–6′*) with microsecond resolution. The prorous silicon was obtained using a current $J=$ 70 mA/cm$^2$ and an etch time of 5 minutes. The sensitivity of curves *1′,2′,3′,4′,5′,6′* is increased by a factor of 2. Delay times, $\mu$s: *1* — 0, *2* — 5, *3* — 10, *4* — 15, *5* — 20, *6* — 30.

trum of long-lived PL is shifted towards the short-wavelength region of the spectrum by roughly 50 nm.

The observed changes in the spectra of long-lived PL from porous silicon when a diamondlike film is deposited are also probably connected with the formation of carbon nanoclusters on the surface of the silicon filaments. This conclusion is confirmed by measurements of the PL made for porous silicon layers obtained at different current densities and consequently having differing porosities.[4] Since a smaller number of carbon atoms will penetrate into the pores during the process of diamondlike film deposition when the porosity decreases, it is logical to assume that in this case the number and possibly the size of the carbon clusters adsorbed at the surface of the silicon filaments will also become smaller. In fact, it follows from Figs. 3 and 4, which show PL spectra (taken with microsecond resolution) of porous silicon obtained for $J=35$ and 50 mA/cm$^2$ respectively, that there is no significant shift in the 660 nm band when the diamondlike film is deposited, in contrast to porous silicon samples with higher porosity (Fig. 2, $J=$ 75 mA/cm$^2$). In this case the intensity of the corresponding bands is somewhat lower than that of the original porous silicon. It should be noted that the observed changes in the PL spectrum of porous silicon plus diamondlike film also can be related to the formation of carbon cluster–silicon chemical bonds in the process of carbon adsorption at the boundaries, and consequently to the appearance of new centers for radiative recombination. The latter can have a particularly strong effect on the PL intensity of the porous silicon-plus-diamondlike film system, depending on the porosity of the original porous silicon, which is also observed in experiment (Figs. 3 and 4). On the other hand, the changes in the PL spectrum described here when diamondlike films are deposited cannot be explained by absorption of light in the diamondlike film, nor as manifestations of features in the absorption spectrum of the diamondlike film. The spectral dependence shown in Fig. 5 for the absorption coefficient of a diamondlike film shows that it is highly transprent in the spectral region under study (out to 300 nm) and that its spectrum is featureless. Note also that the optical width of the band gap of the diamondlike films used was about 4 eV.

Because of the low deposition temperatures and high growth rates of the diamondlike films[11] used in our experiments, the observed changes in the luminescence properties of the porous silicon cannot be explained by removal of adsorbates from the surface of the silicon wires either,[8] nor by surface modification in the rf discharge plasma, not to mention changes in the structure of the silicon nanocrystallites. The effects described above are in our view caused by formation of carbon clusters at the surface of the silicon fila-

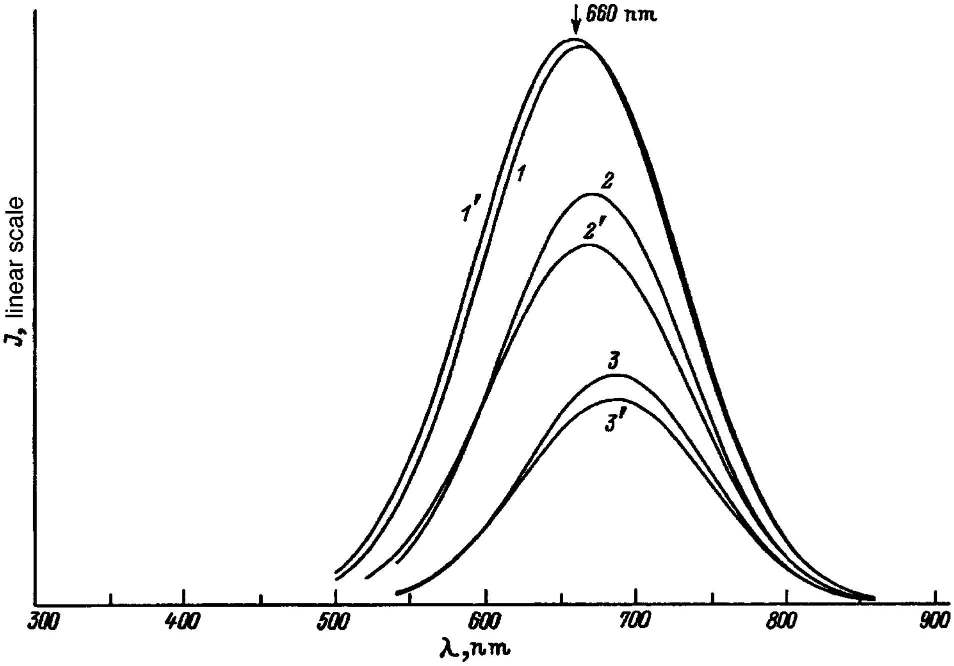

FIG. 3. The same as in Fig. 2 but for porous silicon obtained using a current $J=$ 35 mA/cm$^2$ and an etch time of 10 minutes. *1–3* — porous silicon, *1′–3′* — porous silicon plus diamondlike film. The sensitiity of curves *1′–3′* is increased by a factor of 5. Delay time, $\mu$s: *1* — 0, *2* — 5, *3* — 15.

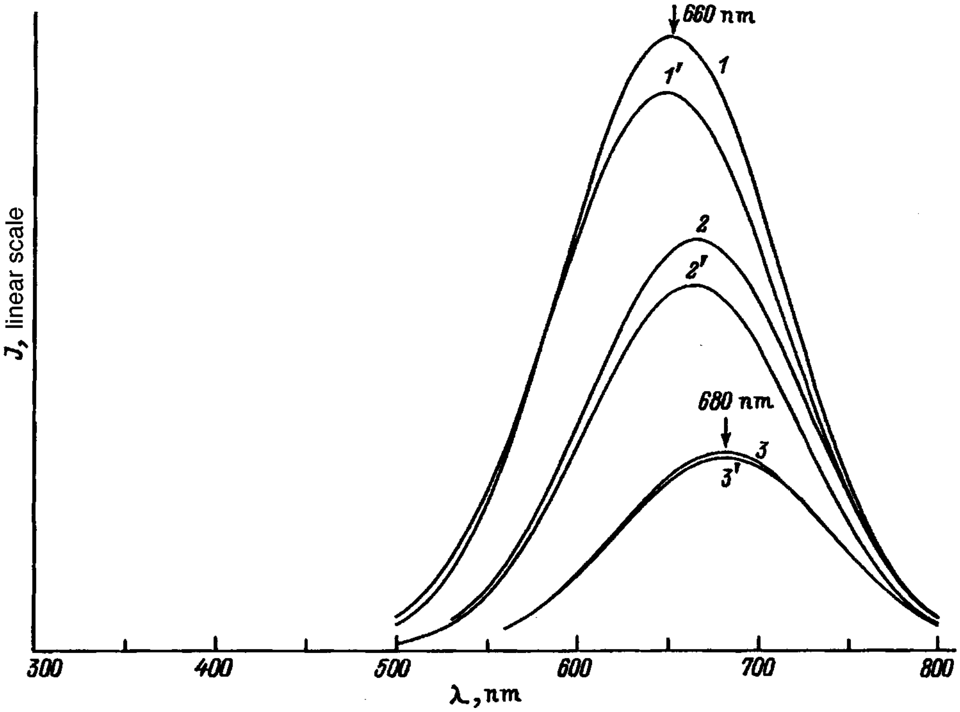

FIG. 4. The same as in Fig. 2, but with porous silicon obtained using a current $J=$ 50 mA/cm$^2$ and an etch time of 5 min. Delay times the same as in Fig. 3.

ments in the process of deposition of the diamondlike films. The presence of clusters, and also the fact that they can act as effective centers for radiative recombination can also lead, on the one hand, to spectral changes in the photoluminescence, particularly in the long-wavelength region of the spectrum, and, on the other hand, to an increase in the intensity of the short-wavelength (nanosecond) photoluminescence.

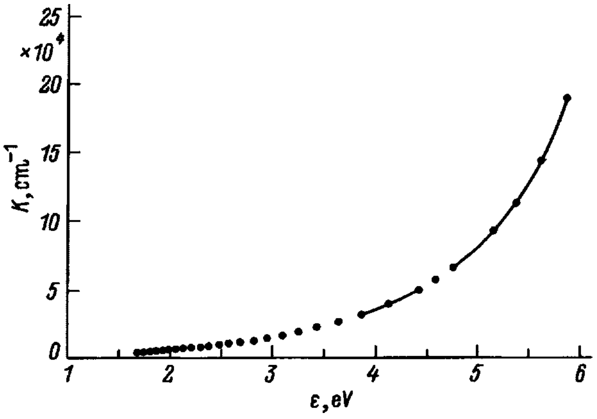

FIG. 5. Spectral dependence of the absorption coefficient $K$ of a diamond-like film ($d=70$ nm) ($\varepsilon$ is the photon energy).

[1] L. T. Canham, Appl. Phys. Lett. **57**, 1046 (1990).

[2] A. G. Cullis and L. T. Canham, Nature **353**, 335 (1991).

[3] A. Halimaoni, C. Oules, G. Bomchil *et al.*, Appl. Phys. Lett. **59**, 304 (1991).

[4] S. V. Svechnikov, A. V. Sachenko, G. A. Sukach *et al.*, Optoėktron. Poluprovodn. Tekh. **27**, 3 (1994).

[5] M. Stutzman, M. S. Brandt, M. Rosenbauer *et al.*, Phys. Rev. **47**, 4806 (1993).

[6] M. S. Brandt *et al.*, Solid State Commun. **8**, 307 (1991).

[7] F. Koch, V. Petrova-Koch, T. Muschik *et al.*, Mater. Res. Soc. Symp. Proc. **238**, 137 (1993).

[8] K. N. El'tsov, V. A. Karavanskiĭ, and V. V. Martynov, JETP Lett. **63**, 119 (1996).

[9] Feng Yan, Xi-mao Bao, and Xiao-wei Wu, Appl. Phys. Lett. **67**, 3471 (1995).

[10] Yu. P. Piryatinskii and M. V. Kurik, Mol. Mater. **43**, 1 (1992).

[11] V. A. Semenovich and N. I. Klyui, J. CVD **4**, 29 (1995).

[12] I. Watanabe and M. Inoue, J. Appl. Phys. **22**, L176 (1983).

[13] A. A. Babaev, M. Sh. Abdulvagabov, E. I. Terukov *et al.*, Neorg. Mater. **27**, 2205 (1991).

Translated by Frank J. Crowne

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Solution of dispersion relations for planar waveguides in the case of complex roots

A. A. Romanenko and A. B. Sotskiĭ

*Institute of Applied Optics, Academy of Sciences of Belarus, 212793 Mogilev, Belarus*

(Submitted April 30, 1996)
Zh. Tekh. Fiz. **68**, 88–95 (April 1998)

A method for calculating the complex roots of a nonlinear equation is described whereby the solution of the problem is reduced to quadratures. Applications of the method to the investigation of dispersion relations for various open waveguide structures with a complex dielectric permittivity are discussed. The possibilities of the prismatic excitation of modes corresponding to the roots of the dispersion relations on different Riemann sheets are analyzed. Solutions are obtained for the inverse problems of reconstructing complex mode propagation constants and determining the parameters of films that guide waveguide and leaky modes. The solution is based on processing of the angular dependence of the reflection coefficient in a prismatic excitation scheme. *© 1998 American Institute of Physics.* [S1063-7842(98)01604-3]

## INTRODUCTION

The rigorous determination of complex roots of dispersion relations is of fundamental importance in the electrodynamic theory of open-ended waveguides. Despite the long history of this problem, a satisfactory solution has yet to be found for it. This dilemma accounts for the several alternative computational approaches in use, including various interpolation[1–4] and gradient[5,6] methods. Their common shortcoming is the need to specify a sufficiently accurate zeroth approximation for the root and to calculate the derivatives of the dispersion relations, which poses a rather complex problem. Moreover, smoothness of the functions involved in the equations is essential for convergence, but unfortunately they suffer discontinuities at branch cuts. Another fundamental requirement is nondegeneracy of the roots.

Here we propose a method, free of these limitations, for rigorously calculating the complex roots of dispersion relations for planar waveguides having an arbitrary distribution of the complex dielectric permittivity. The method is an elaboration of previously published results.[7] It is based on a contour integration technique and can be used to find all the roots of a nonlinear equation $f(u)=0$ in the domain of analyticity of the function $f(u)$.

## 1. METHOD FOR CALCULATING THE COMPLEX ROOTS OF NONLINEAR EQUATIONS

Let us suppose that it is required to find the roots of the equation $f(u)=0$ in a simply connected, closed domain $G$ of the complex variable $u$ bounded by the contour $C$. We also assume that the function $f(u)$ is analytic in this domain. The total number of roots $m$ (taking their multiplicity into account) can be determined on the basis of the argument principle,[8] whereby $m$ is equal to $1/2\pi$ times the total variation of the argument of the quantity $W=f(u)$ in traversing the contour $C$. We assume initially that all the roots are nondegenerate. If $m=1$, then by calculating the integrals $I_c^{(0)}$ and $I_c^{(1)}$ numerically, where

$$I_c^{(k)}=\frac{1}{2\pi i}\oint_c\frac{u^k du}{f(u)},$$

and invoking the residue theorem, we find the value of the root $u_1=I_c^{(1)}/I_c^{(0)}$. If $m>1$, then by successively shrinking the domain and calculating the variations of the argument of $W$, we obtain a domain $G_1$ that is bounded by a contour $C_1$ and contains $m-1$ roots. For the excluded root we obtain

$$u_1=\frac{I_{c_1}^{(1)}-I_c^{(1)}}{I_{c_1}^{(0)}-I_c^{(0)}}.$$

Repeating the process, we find all the roots in succession. We note that the calculation of the variation of the argument of $W$ reduces to the calculation of the number of crossings of the boundaries of the coordinate quadrants. This operation can be executed in parallel with the accumulation of integral sums, enhancing the computational efficiency of the method.

We now generalize the computational scheme to the case involving a double root $u_0$. We treat this situation as the limiting case when $u_1\rightarrow u_2\rightarrow u_0$, where $u_1$ and $u_2$ are simple roots. We assume that two roots are present in the domain $G$. We introduce integrals $I_c^{(2)}$ and $I_c^{(3)}$ in addition to $I_c^{(0)}$ and $I_c^{(1)}$. Using the residue theorem, we arrive at the system of four equations

$$I_c^{(k)}=u_1^k[f'(u_1)]^{-1}+u_2^k[f'(u_2)]^{-1}\quad(k=0,1,2,3)\qquad(1)$$

in the four unknowns $u_1$, $u_2$, $f'(u_1)$, and $f'(u_2)$. Its solution has the form

$$u_{1,2}=a/2\pm\sqrt{(a/2)^2-b},\qquad(2)$$

$$a=\frac{I_c^{(1)}I_c^{(2)}-I_c^{(0)}I_c^{(3)}}{[I_c^{(1)}]^2-I_c^{(0)}I_c^{(2)}},\qquad(3)$$

$$b=\frac{[I_c^{(2)}]^2-I_c^{(3)}I_c^{(1)}}{[I_c^{(1)}]^2-I_c^{(0)}I_c^{(2)}}. \tag{4}$$

Forming Taylor expansions of the function $f(u)$ and its derivatives, we obtain

$$f'(u_1)=-\frac{1}{2}f''(u_1)\Delta u-\frac{1}{6}f'''(u_1)(\Delta u)^2+O[(\Delta u)^3], \tag{5}$$

$$f'(u_2)=\frac{1}{2}f''(u_1)\Delta u+\frac{1}{6}f'''(u_1)(\Delta u)^2+O[(\Delta u)^3], \tag{6}$$

$$f''(u_2)=f''(u_1)+f'''(u_1)\Delta u+O[(\Delta u)^2], \tag{7}$$

where $\Delta u=u_2-u_1$. The application of Eqs. (5)–(7) reduces Eqs. (1) to the form

$$I_c^{(0)}=-\frac{2}{3}\frac{f'''(u_1)}{f''(u_1)}+O(\Delta u),$$

$$I_c^{(1)}=\frac{2}{f''(u_1)}+u_1I_c^{(0)}+O(\Delta u),$$

$$I_c^{(2)}=\frac{4u_1}{f''(u_1)}+u_1^2I_c^{(0)}+O(\Delta u),$$

$$I_c^{(3)}=\frac{6u_1^2}{f''(u_1)}+u_1^3I_c^{(0)}+O(\Delta u).$$

It follows from these expressions that $a^2/4\to b$ and $u_1\to u_2\to u_0=a/2$ in the limit $\Delta u\to 0$. For $m>2$ it is necessary once again to shrink the domain. If $m-2$ roots are contained in the domain $G_1$, the values of the two excluded roots can be calculated from Eqs. (2)–(4) after the substitution $I_c^{(k)}\to I_c^{(k)}-I_{c_1}^{(k)}$. For the double root we again obtain $u_0=a/2$.

All not more than twofold-degenerate roots of the equation $f(u)=0$ can be found by combining the above-described computational schemes. The case of $l$-fold degeneracy $(l>2)$ can be treated analogously by computing the integrals $I_c^{(k)}$, where $k=0,1,\ldots,2l-1$. The corresponding expressions are rather cumbersome and will not be written out here, particularly in view of the fact that the roots of the dispersion relations for planar waveguides are not degenerate as a rule, and only in rare situations encountered in the investigation of anisotropic waveguides and systems of coupled waveguides are they twofold degenerate.

## 2. SOLUTION OF THE DISPERSION RELATIONS

We consider a waveguide formed by a layered medium, which has a complex dielectric permittivity, is contained in the domain $-d\leqslant y\leqslant 0$, and is surrounded by homogeneous media with dielectric constants $\varepsilon_g(y>0)$ and $\varepsilon_s(y<-d)$. The dispersion relation for modes whose fields are exponential functions of the time and the coordinate $z$, $\exp(i\omega t-ihz)$ has the form[7,9]

$$F_\nu=i\psi(0)\nu+\psi'(+0)=0, \tag{8}$$

where $\psi(y)$ has the meaning of the component $E_x$ for the TE modes and $H_x$ for the TM modes, $\nu=\sqrt{k_0^2\varepsilon_g-h^2}$, and $k_0=2\pi/\lambda_0$ is the wave number in vacuum.

In the derivation of Eq. (8) we have chosen the functional dependence $\psi(y)=\psi(0)\exp(-i\nu y)$, $y\geqslant 0$. We use a stratification method to specify the quantities appearing in Eq. (8), representing the waveguide by a set of $n$ homogeneous layers.[10] In this case the quantities $\psi'(+0)$ and $\psi(0)$ can be calculated from the recursion relations[7]

$$\psi_{j+1}=\psi_jK+\psi_j'S, \tag{9}$$

$$\psi_{j+1}'=(\varepsilon_{j+1}/\varepsilon_j)^T(\psi_j'K-\psi_j\nu_j^2S), \tag{10}$$

$$\psi_1=1, \tag{11}$$

$$\psi_1'=i\nu_1(\varepsilon_2/\varepsilon_1)^T, \tag{12}$$

where $S=\sin(\nu_j\Delta y_j)/\nu_j$, $K=\cos(\nu_j\Delta y_j)$, $\nu_j=\sqrt{k^2\varepsilon_j-h^2}$, $\varepsilon_1=\varepsilon_s$, $\varepsilon_{n+2}=\varepsilon_g$, $\psi_{n+2}=\psi(0)$, $\psi_{n+2}'=\psi'(+0)$, $\varepsilon_j$ and $\Delta y_j$ are the permittivity and thickness of the $j$th layer, $T=0$ for TE modes, and $T=1$ for TM modes; the field in the domain $y<-d$ is represented as $\psi(y)=\exp[i\nu_i(y+d)]$.

In the special case of a homogeneous thin-film waveguide $(n=1)$, Eq. (8) with allowance for Eqs. (9)–(12) reduces to the form

$$F_\nu=\left[\nu_1\left(\frac{\varepsilon_2}{\varepsilon_s}\right)^T+\nu\left(\frac{\varepsilon_2}{\varepsilon_g}\right)^T\right]\cos(\nu_2d)+i\left[\nu_2+\frac{\nu_1\nu}{\nu_2}\left(\frac{\varepsilon_2^2}{\varepsilon_s\varepsilon_g}\right)^T\right]\sin(\nu_2d)=0. \tag{13}$$

It is convenient to choose $u=\nu_1$ as the unknown in Eqs. (8) and (13). Then $\nu_j=\sqrt{k_0^2(\varepsilon_j-\varepsilon_s)+u^2}$ $(j=2,\ldots,n+2;$ $\nu_{n+2}=\nu)$. According to Eqs. (8)–(13), the function $F_\nu(u)$ is bounded (in a finite part of the complex plane) and invariant with respect to the choice of signs of $\nu_j$ $(j<n+2)$; the only source of nonanalyticity of the function is the presence of branch points and branch cuts of the function $\nu(u)$. The nonanalyticity is eliminated by working with the product

$$f(u)=F_\nu(u)F_{-\nu}(u), \tag{14}$$

which is an entire function of the variable $u$ (Ref. 9). The roots of Eq. (14) coincide with those of Eqs. (8) and (13) corresponding to two branches of the function $\nu(u)$ (e.g., the branches Im $\nu\leqslant 0$ and Im $\nu\geqslant 0$). We note that for $\nu\neq 0$ Eq. (14) and the equation $F_{\pm\nu}(u)=0$ have roots of identical multiplicity. Indeed, the multiplicity of the roots increases if the system of equations $F_\nu(u)=0$, $F_{-\nu}(u)=0$ holds, from which, according to (8), it follows that $\psi(0)=0$ and $\psi'(+0)=0$ $(\nu\neq 0)$. In this case the solution of the Cauchy problem for the differential equation describing the mode field gives $\psi(y)\equiv 0$ and $\psi'(y)\equiv 0$, contradicting conditions (11) and (12). The case $\nu\neq 0$, on the other hand, can be analyzed separately. We also note the implication of Eqs. (8)–(12), that if $d\neq 0$, then the quantity $|W|$ grows exponentially as $|u|\to\infty$. In this case the number of roots of Eq. (14) is $m=O(r)$ in the limit $r\to\infty$, where $r$ is the radius of the circle $C$ (Ref. 9), i.e., when $G$ is interpreted as the entire complex plane, the number $m$ is inboundedly large. But if

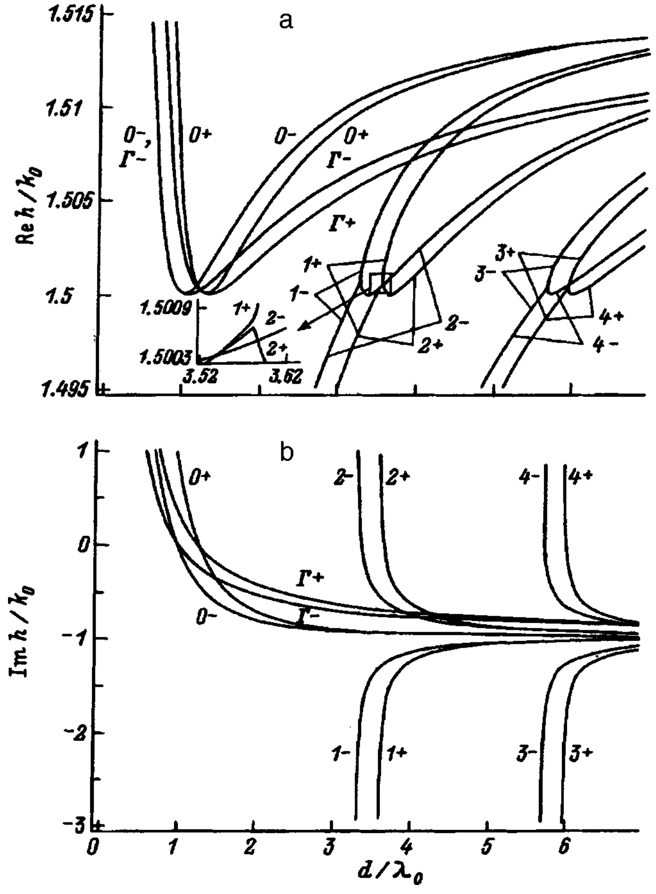

FIG. 1. Dispersion curves for the modes of a homogeneous waveguide and an inhomogeneous waveguide. a) Re $h/k_0$; b) Im $h/k_0$.

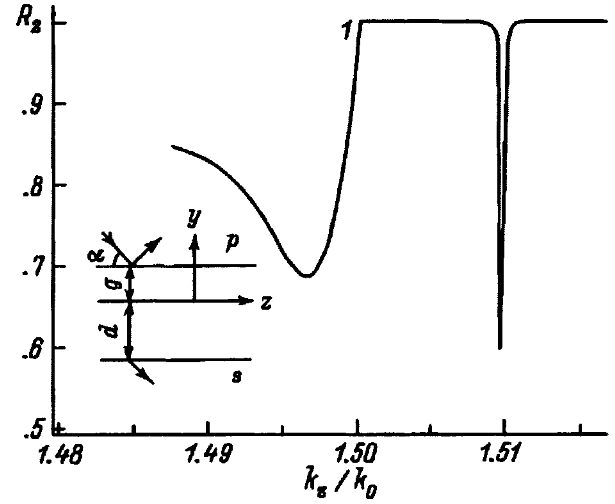

FIG. 2. Schematic diagram of the prismatic excitation device and graph of the reflection coefficient of the exciting wave versus the $z$-component of its wave vector for $\varepsilon_p=3.811$.

$d=0$, we infer from (13) that $m=2$ for $T=1$, $m=0$ for $T=0$ and $\varepsilon_g\neq\varepsilon_s$, and $m=1$ for $T=0$ and $\varepsilon_g=\varepsilon_s$.

Using the method described in Sec. I, we have obtained solutions of Eq. (14) for several waveguide structures. Typical plots of $h=\sqrt{k_0^2\varepsilon_s-\nu_1^2}$ (Re $h\geqslant 0$) as a function of $d/\lambda_0$ are shown in Figs. 1a and 1b. They have been obtained for TE-polarized modes of a homogeneous waveguide with $\varepsilon_s=2.25-i3\times10^{-6}$, $\varepsilon_2=2.295225-i3.03\times10^{-6}$, and $\varepsilon_g=1$. The "+" and "−" signs in the figures identify curves pertaining to Im $\nu\geqslant 0$ and Im $\nu\leqslant 0$, respectively. Curves *1*± and *3*± correspond to Im $\nu_1>0$, Re $\nu_1>0$. Curves *1*+ and *3*+, for which Re $\nu>0$, correspond to leaky waves exiting from the waveguide into the two open domains $y>0$ and $y<-d$, while curves *1*− and *3*− (Re $\nu<0$) correspond to waves which are leaky only in the domain $y<-d$. The inequality Re $\nu_1<0$ holds for curves *0*±, *2*±, and *4*±. The indicated curves in Fig. 1a have points of tangency with the line Re $h/k_0=1.5$. These points correspond to critical thicknesses $d=d_k$ ($k$ denotes the symbol enumerating the curves). For $d>d_k$ we have Re $\nu<0$ and Im $\nu_1>0$ ($k=0-,2-,4-$) or Re $\nu>0$ and Im$p_1<0$ ($k=0+,2+,4+$). In this case curves *0*−, *2*−, and *4*− describe the usual dispersion curves for waveguide modes, whereas curves *0*+, *2*+, and *4*+ correspond to waves leaking from the waveguide into the domain $y>0$. For $d<d_k$ we have Re $\nu>0$ and Im $\nu_1>0$ ($k=0-,2-,4-$) or Re $\nu<0$ and Re $\nu>0$ ($k=0+,2+,4+$). In this case curves *0*−, *2*−, and *4*− correspond to waves entering the waveguide from the domain $y<-d$ (they grow as $y\to-\infty$), then leaking from the waveguide into the domain $y>0$ and decaying as $y\to\infty$, whereas curves *0*+, *2*+, and *4*+ correspond to waves entering the waveguide from both open domains and growing as $|y|\to\infty$. A distinctive feature of all these waves is the growth of the field as $z\to\infty$ (Im $h>0$, Fig. 1b). It follows from the above discussion that the sets of curves (*1*±, *2*±) and (*3*±, *4*±) correspond to modes having similar properties. The solution of Eq. (14) has shown that increasing $d$ generates new sets (*5*±, *6*±),(*7*±, *8*±),..., analogous to those already mentioned. The indicated prominent characteristics of the behavior of the dispersion relations and the mode fields remain in effect for TM-polarized modes. And they are similar in regard to inhomogeneous waveguides. This statement is illustrated by curves Γ+ and Γ− in Figs. 1a and 1b, which are calculated for TE modes of a waveguide having the index profile $\varepsilon(y)=2.25-i3\times10^{-6}+(0.045225-i2.73\times10^{-5})\exp[-(y/d)^2]$, $y\leqslant 0$; $\varepsilon(y)=\varepsilon_g=1$, $y>0$. These curves are analogous to curves *0*+ and *0*− discussed above. We note, however, that the results here are characteristic of waveguides having a higher permittivity than the surrounding media. The dispersion relations for waveguide media having a lower permittivity (e.g., metal films on dielectric substrates and low-refraction films on high-refraction substrates) exhibit distinct differences. We shall examine them briefly below. We also note that the above-discussed inequalities Im $\nu_1>0$ and Im $\nu>0$ correspond to modes whose fields grow exponentially with increasing distance from the waveguide. Such modes do not occur in the complete sets of modes,[9] making it necessary to ascertain the possibilities of their excitation and practical utilization.

We consider the excitation of modes by means of a prismatic coupler.[11,12] A schematic diagram of the coupler is shown in Fig. 2. The prism P (which we consider to be unbounded for simplicity) is separated from the waveguide by a buffer layer of thickness $g$. The prism and the buffer layer have real permittivities $\varepsilon_p$ and $\varepsilon_g$, where $\varepsilon_g < \varepsilon_p$. The structure is excited by a plane wave whose wave vector forms an angle $\alpha$ with the base of the prism. Examining the recursion relations (9) and (10) in the buffer layer, we obtain the following expression for the reflection coefficient of the exciting wave:

$$R=\frac{(1-\delta)F_\nu-\exp(-2i\nu g)(1+\delta)F_{-\nu}}{(1+\delta)F_\nu-\exp(-2i\nu g)(1-\delta)F_{-\nu}},$$

$$\delta=(\nu/\nu_p)(\varepsilon_p/\varepsilon_g)^T,\quad \nu=\sqrt{k_0^2\varepsilon_g-k_z^2},\quad \mathrm{Im}\,\nu\leqslant 0,$$

$$\nu_p=\sqrt{k_0^2\varepsilon_p-k_z^2},\ \ \mathrm{Re}\,\nu_p\geqslant 0,\ \ k_z=k_0\sqrt{\varepsilon_p}\cos\alpha. \tag{15}$$

The quantities $F_\nu$ and $F_{-\nu}$ can be calculated from Eqs. (8)–(13), where $\nu_j=\sqrt{k_0^2\varepsilon_j-k_z^2}$ and in accordance with the radiation condition $\mathrm{Re}\,\nu_1\geqslant 0$.

We know that the efficient excitation of waveguide modes characterized by the occurrence of resonance troughs in the $R_2(k_z)$ curve ($R_2=|R|^2$) takes place under the conditions of weak prism–waveguide coupling:[11,12]

$$k_z^2>k_0^2\varepsilon_g, \tag{16}$$

$$\exp(-i\nu g)\ll 1. \tag{17}$$

According to Eqs. (8)–(13) and (15)–(17), $R_2$ can differ significantly from unity only if $F_\nu=O[\exp(-2i\nu g)]$. Making use of the fact that the variable $k_z$ in Eq. (15) belongs to the real axis of the Riemann sheet $\mathrm{Re}\,\nu_1>0$, we infer that the stated condition can be satisfied if there is a domain $H$, defined by the inequality $|k_z-h|<\rho$ ($h$ is the propagation constant of the excited mode, and $\rho=O[\exp(-2i\nu g)]$), which contains a segment of the real axis of the sheet $\mathrm{Re}\,\nu_1>0$ and in which the function $F_\nu(k_z)$ is analytic. These considerations are consistent with the obvious fact that resonance excitation is admissible only for modes whose fields decay with increasing distance from the waveguide in the buffer layer and whose propagation constants satisfy the equation $F_\nu(h)=0$.

The analyticity properties of the function $F_\nu(k_z)$ in the vicinity of the roots can be assessed from Figs. 3a and 3b. The solid curves represent the function $\mathrm{Im}\,k_z(\mathrm{Re}\,k_z)$ [equivalent to $\mathrm{Im}\,h(\mathrm{Re}\,h)$] for the roots of the equation $F_\nu(k_z)=0$. The dashed lines indicate the cuts for the function $\nu_1(k_z)$ as defined by the conditions $\mathrm{Im}\,k_z=k_0^2\,\mathrm{Im}\,\varepsilon_s/(2\,\mathrm{Re}\,k_z)$ and $k_0^2\,\mathrm{Re}\,\varepsilon_s-(\mathrm{Re}\,k_z)^2+(\mathrm{Im}\,k_z)^2<0$. The cuts pass below the real axes $\mathrm{Im}\,k_z=0$. Curves *0−* through *4−* correspond to their counterparts in Figs. 1a and 1b. It follows from Fig. 3a that only the segments of curves *1−* and *3−* to the left of the branch point of the function $\nu_1(k_z)$ can exist in the domain $H$. An analogous situation is met for the segments of curves *0−*, *2−*, and *4−* below the cut. In this case the domain $H$ belongs to a two-sheeted Riemann surface "glued" along the edges of the cut. Now, looking at Figs. 1a and 1b, we infer that the efficient excitation of waveguide modes is possible only for waveguide films whose thicknesses exceed critical values ($d>d_k$). On the other hand, resonance troughs corresponding to the excitation of leaky modes associated with curves *1−*, *3−* and their analogs *5−*, *7−*,... can be observed for $d<d_k$. These conclusions are illustrated in Fig. 2, curve *1*, which is calculated on the basis of Eqs. (9)–(12) and (15) for $d/\lambda_0=3$ and $g/\lambda_0=0.16$. Its right minimum corresponds to a waveguide (*0−*) mode, and its left minimum corresponds to a leaky (*1−*) mode.

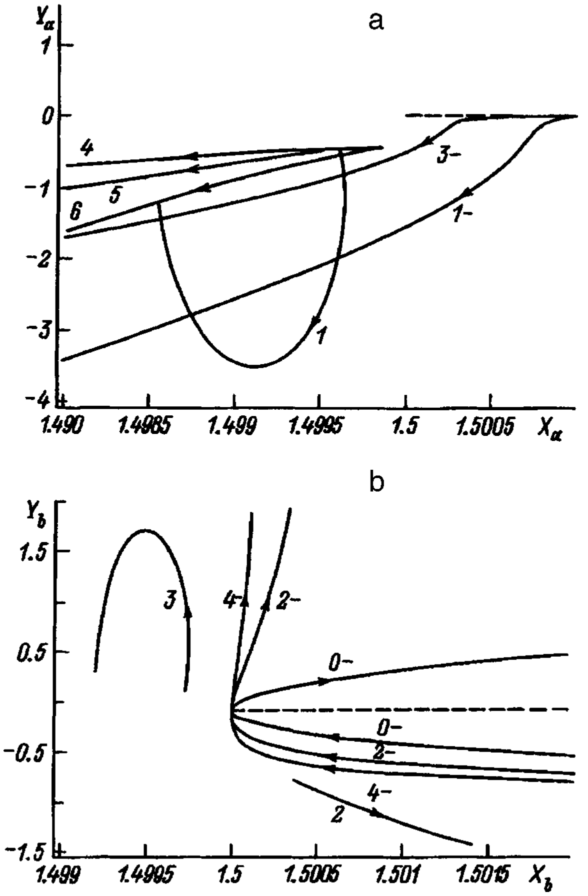

FIG. 3. Positions of the roots of the equation $F_\nu(k_z)=0$ on the Riemann sheets $\mathrm{Re}\,\nu_1>0$ (a) and $\mathrm{Re}\,\nu_1<0$ (b). $X_{a,b}=A_{a,b}+B_{a,b}\,\mathrm{Re}\,k_z/k_0$, $Y_{a,b}=C_{a,b}+D_{a,b}\,\mathrm{Im}\,k_z/k_0$; for curves *0−* to *4−*: $A_{a,b}=0$, $B_{a,b}=1$, $C_{a,b}=0$, $D_a=10^3$, $D_b=10^5$; *1)* $A_a=1.494$, $B_a=5.45\times10^{-3}$, $C_a=-0.44$, $D_a=30$; *2)* $A_b=1.498$, $B_b=1.47\times10^{-3}$, $C_b=-0.72$, $D_b=19.7$; *3)* $A_b=1.497$, $B_b=2.65\times10^{-3}$, $C_b=-9.77\times10^{-2}$, $D_b=14.8$; *4–6)* $A_a=4.35\times10^{-2}$, $B_a=1$, $C_a=0$, $D_a=4\times10^4$. The arrows indicate directions of decreasing $d$.

We now discuss the prismatic mode excitation of thin films having a lower real part of the permittivity than the surrounding media. Curves *1–3* in Fig. 3a and 3b represent solutions of the equation $F_\nu(k_z)=0$ for optical-range TM modes guided by a silver film ($\varepsilon_2=-18-i0.47$) surrounded by dielectric media ($\varepsilon_g=1\varepsilon_s=2.25-i3\times10^{-6}$). On curve *1* we have $\mathrm{Re}\,\nu_1>0$ and $\mathrm{Im}\,\nu_1>0$, i.e., it refers to waves leaking into the domain $y<-d$. The start of the curve ($d\to\infty$) corresponds to a plasmon mode of the interface $y=0$, and the end ($d\to 0$) corresponds to a surface-wave mode of the interface between two dielectrics. Curve *2*, along which $\mathrm{Re}\,\nu_1<0$ and $\mathrm{Im}\,\nu_1<0$, begins at the point corresponding to a plasmon mode of the interface $y=-d$ and goes to infinity [$\mathrm{Re}(k_z/k_0)\to\infty$, $\mathrm{Im}(k_z/k_0)\to-\infty$] as $d\to 0$. On curve *1* we have $\mathrm{Re}\,\nu_1<0$ and $\mathrm{Im}\,\nu_1>0$. It refers to waves entering the waveguide from the domain $y<-d$, their fields growing as $y\to-\infty$ and $z\to\infty$. According to the preceding discussion, modes corresponding to curves *1* and *2* admit resonance excitation. An additional restriction for such excitation is given by condition (16), which, in particular, rules out the possibility of the excitation of a dielectric-dielectric interface (surface-wave) mode. Curves *4–6* in Fig. 3a represent solutions of the equation $F_\nu(k_z)=0$ for the three lowest optical-range TE modes guided by a silicon dioxide film ($\varepsilon_2=2.121975-i2.9134\times10^{-6}$) terminated in air ($\varepsilon_g=1$) and in a silicon substrate ($\varepsilon_s=15.21-i0.39$). For all these curves we have Re $\nu_1>0$ and Im $\nu_1>0$, i.e., they refer to modes leaking into the substrate. Curves *4–6* begin at the point $k_z/k_0=\sqrt{\varepsilon_2}$ $(d\rightarrow0)$, which is far from the cut of the function $\nu_1(k_z)$ (outside the field of view of the figure). Accordingly, all the investigated modes admit resonance excitation. We have confirmed these conclusions by calculations of the $R_2(k_z)$ curves, which are characterized by resonance troughs (similar to those in Fig. 2) corresponding to excitation of the indicated modes.

## 3. INVERSE PROBLEMS OF RECONSTRUCTING COMPLEX MODE PROPAGATION CONSTANTS AND THE PARAMETERS OF WAVEGUIDE FILMS

The measurement of the complex mode propagation constants $h$ plays an important role in the investigation of the properties of waveguides and surface layers. An approach based on experimental recording of the resonance troughs of the $R_2(k_z)$ curves[13–16] is widely used at the present time to determine $h$. In this approach, however, rather limited information about the indicated curves is taken into account; only the coordinates of the minima of the functions $R_2(k_z)$ are measured, and they are identified with the real parts of the propagation constants Re $h$. The imaginary parts Im $h$ are determined in additional measurements, where attenuation of the modes is observed along the direction of their propagation,[14] This procedure for the determination of $h$ is time-consuming in the experimental realm and has a fundamental shortcoming in that the perturbing influence of the prism on the investigated structure is ignored.

One of the important applications of waveguide methods is to reconstruct the parameters of waveguide films In this case the values of $h$ for two modes of known order can be used to write a system of two dispersion relations containing inverse trigonometric functions; the numerical solution of these equations gives the quantities $\varepsilon_2$ and $d$ (Refs. 13 and 14). However, difficulties are encountered in connection with the rigorous solution of the system, owing to the presence of branch points and cuts for the inverse trigonometric functions. This problem is particularly significant when modes existing under near-critical conditions are used.

More efficient approaches to the solution of the indicated inverse problems are described below. The real and imaginary parts of a mode propagation constant of a planar waveguide structure (inhomogeneous in the general case) are determined simultaneously with allowance for the perturbing influence of the prism by integrating the function $R_2(k_z)$ in the vicinity of the resonance troughs. The quantities $\varepsilon_2$ and $d$ are determined by the system of equations (13) for two values of $h$. The solution of the system reduces to the determination of the roots of entire functions by the method discussed above.

Expanding the function $F_\nu(k_z)$ in a Taylor series in the domain $H$, we obtain the following approximation for the function $R(k_z)$ in the vicinity of resonances[7]:

$$R=\frac{1-\delta}{1+\delta}-\frac{4\delta}{1-\delta^2}\,\frac{\Delta h}{k_z-\overline{h}}+0(|\Delta h|). \tag{18}$$

Here $\delta$ is calculated for $k_z=\mathrm{Re}\,h$, $\overline{h}$ is the mode propagation constant of the waveguide–prism system, and $\Delta h=\overline{h}-h$. Equation (18) is defined in the case of the excitation of an arbitrary plane-layered waveguide structure, for which

$$\Delta h=-\frac{i\nu(1-\delta)\exp(-2i\nu g)}{(\varepsilon_g)^T(1+\delta)I\ \mathrm{Re}\,h}, \tag{19}$$

$$I=\int_{-d}^{\infty}\frac{\psi^2(y)}{[\varepsilon(y)]^T}\,dy+\frac{\psi^2(-d)}{2i\nu_1(\varepsilon_s)^T}, \tag{20}$$

where the function $\psi(y)$ with the normalization $\psi(0)=1$ describes the distribution of the field of the excited mode, and

$$\nu_1=\sqrt{k_0^2\varepsilon_s-h^2},\quad \nu=-i\sqrt{(\mathrm{Re}\,h)^2-k_0^2\varepsilon_g}.$$

To solve the inverse problem for $h$, it is convenient to introduce the parameters[17]

$$p_1=\frac{\mathrm{Im}\,\overline{h}}{k_0}, \tag{21}$$

$$p_2=-\frac{2\delta\nu\exp(-2i\nu g)}{(\varepsilon_g)^T(1-\delta)Ik_0\ \mathrm{Re}\,h}, \tag{22}$$

$$p_4=\frac{\mathrm{Re}\,\overline{h}-k_z}{k_0}, \tag{23}$$

which characterize the mode attenuation of the external prismatic structure, the coupling of the prism with the waveguide, and the deviation of the $z$-component of the exciting wave vector from the resonance value of Re $\overline{h}$. The parameter $p_2$ is complex-valued in general, i.e., $p_2=|p_2|\exp(i\sigma)$. However, in the case of primary practical interest—weakly damped modes—the inequalities $|\mathrm{Re}\,I|\gg|\mathrm{Im}\,I|$ and $\sigma\ll1$ hold. They permit small quantities $O(\sigma^2)$ to be omitted in the subsequent calculations. Taking Eqs. (18)–(23) into account, we obtain

$$R_2=|R|^2=1+[4|p_2|(p_1+|p_2|-p_4\sigma)]/(p_1^2+p_4^2). \tag{24}$$

It follows from Eq. (24) that the presence of mode attenuation ($\sigma\neq0$) parts asymmetry to the function $R_2(p_4)$, the degree of asymmetry increasing as the losses increase (Fig. 2). We now determine the coordinate of the minimum of $R_2(p_4)$, denoting it by $p_4^{(0)}$. Differentiating Eq. (24), we obtain

$$p_4^{(0)}=-\frac{p_1^2\sigma}{2(|p_2|+p_1)}, \tag{25}$$

TABLE I.

| Mode order | Polari-zation | $h/k_0$ (exact) | $g/\lambda_0$ | $h/k_0$ (reconstructed) | Mode order | $\varepsilon_2$ (reconstructed) | $d/\lambda_0$ |
|---|---|---|---|---|---|---|---|
| Waveguide film, $\varepsilon_2=2.295225-i3.03\times10^{-5}$, $d/\lambda_0=5$ | | | | | | | |
| 1 | $TE_W$ | $1.512636-i9.819\times10^{-6}$ | 0.28 | $1.512636-i9.819\times10^{-6}$ | 1–2 | $2.295225-i3.030\times10^{-5}$ | 4.99998 |
| 2 | $TE_W$ | $1.505806-i8.999\times10^{-6}$ | - | $1.505805-i8.999\times10^{-6}$ | 1–3 | $2.295222-i3.030\times10^{-5}$ | 5.00089 |
| 3 | $TM_W$ | $1.512560-i9.805\times10^{-6}$ | - | $1.512560-i9.806\times10^{-6}$ | 2–3 | $2.295225-i3.030\times10^{-5}$ | 4.99999 |
| Waveguide film, $\varepsilon_2=2.295225-i3.03\times10^{-5}$, $d/\lambda_0=3$ | | | | | | | |
| 1 | $TE_W$ | $1.509681-i9.295\times10^{-6}$ | 0.16 | $1.509674-i9.296\times10^{-6}$ | 1–2 | $2.295251-i3.031\times10^{-5}$ | 2.9943 |
| 2 | $TE_L$ | $1.498070-i3.363\times10^{-3}$ | - | $1.498177-i3.585\times10^{-3}$ | 1–3 | $2.295090-i3.029\times10^{-5}$ | 3.0131 |
| 3 | $TM_W$ | $1.509436-i9.223\times10^{-6}$ | - | $1.509432-i9.222\times10^{-6}$ | 2–3 | $2.295269-i3.031\times10^{-5}$ | 2.9937 |
| Silver film, $\varepsilon_2=-18-i0.47$, $d/\lambda_0=0.06$ | | | | | | | |
| 1 | $TM_L$ | $1.031124-i4.595\times10^{-3}$ | 1.6 | $1.031198-i4.668\times10^{-3}$ | 1–2 | $-18.0039-i0.4694$ | 0.06014 |
| 2 | $TM_W$ | $1.625144-i4.617\times10^{-3}$ | 0.16 | $1.624916-i4.595\times10^{-3}$ | | | |
| Silicon dioxide, $\varepsilon_2=2.121875-i2.913\times10^{-5}$, $d/\lambda_0=3$ | | | | | | | |
| 1 | $TE_L$ | $1.448039-i2.524\times10^{-4}$ | 0.32 | $1.448039-i2.526\times10^{-4}$ | 1–2 | $2.121977-i2.966\times10^{-5}$ | 2.99981 |
| 2 | $TE_L$ | $1.421781-i9.937\times10^{-4}$ | - | $1.421779-i9.960\times10^{-4}$ | 1–3 | $2.121977-i2.967\times10^{-5}$ | 2.99983 |
| 3 | $TE_L$ | $1.377055-i2.279\times10^{-3}$ | - | $1.377052-i2.290\times10^{-3}$ | 2–3 | $2.121981-i3.517\times10^{-5}$ | 2.99984 |

$$R_2^{(0)}=1+\frac{4|p_2|(p_1+|p_2|)}{p_1^2}, \tag{26}$$

where $R_2^{(0)}=R_2(p_4^{(0)})$.

It follows from Eq. (26) that

$$|p_2|=0.5p_1[\operatorname{sgn}(g/g_0-1)\sqrt{R_2^{(0)}}-1]. \tag{27}$$

Here $g_0$ is the thickness of the buffer layer, for which $R_2^{(0)}=0$. To find the parameters $p_1$ and $\sigma$, we isolate the interval $(p_4^{(0)}-\Delta k_z/k_0, p_4^{(0)}+\Delta k_z/k_0)$ of variation of $p_4$ and form the integrals

$$I_1=\int_{-\Delta k_z/k_0}^{0} R_2(p_4^{(0)}+x)dx,$$

$$I_2=\int_{0}^{\Delta k_z/k_0} R_2(p_4^{(0)}+x)dx,$$

where $R_2(p_4)$ is a function of the form (24).

Calculating the quantities $(I_1+I_2)/2$ and $(I_1-I_2)/2$, we obtain

$$\left[1-\frac{1}{2\Delta k_z}\int_{-\Delta k_z}^{\Delta k_z} R_2(k_z^{(0)}+x)dx\right][1-R_2^{(0)}]^{-1}$$

$$=\frac{p_1k_0}{\Delta k_z}\arctan\left(\frac{\Delta k_z}{p_1k_0}\right), \tag{28}$$

$$\sigma=\left[\int_{-\Delta k_z}^{0} R_2(k_z^{(0)}+x)dx-\int_{0}^{\Delta k_z} R_2(k_z^{(0)}+x)dx\right.$$

$$\{4k_0|p_2|[(1+(k_0p_1/\Delta k_z)^2)^{-1}$$

$$-\ln(1+(\Delta k_z/(k_0p_1))^2)]\}^{-1}, \tag{29}$$

where $R_2(k_z)$ is the experimentally recorded function on the interval $(k_z^{(0)}-\Delta k_z, k_z^{(0)}+\Delta k_z)$, and $k_z^{(0)}$ is the coordinate of the minimum of this function.

Consequently, identifying $R_2^{(0)}$ with the minimum value of the function $R_2(k_z)$, we can determine the values of the parameters $p_1$, $p_2$, and $p_4^{(0)}$. We first solve Eq. (29) for $p_1$; this equation has a single root by virtue of the monotonicity of the function $f(x)=x\arctan(x^{-1})$ in the domain $x\leqslant 0$. The quantities $p_2$ and $p_4^{(0)}$ are then determined from Eqs. (25), (27), and (29) by direct calculation. According to Eqs. (21)–(23), the required mode propagation constant is

$$h=k_z^{(0)}+p_4^{(0)}+i[p_1-p_2(1-\delta)^2(2\delta)^{-1}]. \tag{30}$$

We note that the quantity $\delta$ in (30) depends on $\operatorname{Re} h$. However, since $|p_4^{(0)}|$, $|p_1|$, and $|p_2|$ are small, it can be evaluated for $h=k_z^{(0)}$ and, if necessary, refined by an iterative procedure. We also note that the integration operations in Eqs. (28) and (29) ensure stability of the reconstructed value of $h$ against noise of the function $R_2(k_z)$.

We now address the problem of reconstructing the parameters of a waveguide film $\varepsilon_2$ and $d$. We assume that values of $h$ have been found for two modes, the values of $\varepsilon_g$ and $\varepsilon_s$ are given, $\varepsilon_2$ is situated in the domain $G$ of the complex plane, and $d$ lies in the interval $(d_1,d_2)$. We fix a certain $d$ in this interval and substitute the first value of $h$ into Eq. (13). We see at once that $F_\nu(\varepsilon_2)$ is an entire function, so that its roots in the domain $G$ can be found by the scheme of Sec. I. We note that, in general, there can be several roots of this kind [if $G$ is interpreted as the entire complex plane, the number of such roots can be unbounded by virtue of the exponential asymptotic behavior of $F_\nu(\varepsilon_2)$ in the limit $|\varepsilon_2|\to\infty$; Ref. 9]. We denote the roots so obtained by $\varepsilon_{2j}^{(1)}$ $(j=1,2,\dots)$. Solving Eq. (13) analogously, for the second value of $h$ we have a different set of roots $\varepsilon_{2k}^{(2)}$ $(k=1,2,\dots)$. The quantity $\Delta=\min_{j,k}(\Delta_{jk})$ can be found by direct sequential inspection of the differences $\Delta_{jk}=|\varepsilon_{2j}^{(1)}-\varepsilon_{2k}^{(2)}|$. Now, plotting the function $\Delta(d)$, $d\in(d_1,d_2)$, and determining its minimum, we arrive at the required values of $\varepsilon_2$ and $d$.

Table I gives examples of how the above-described computational scheme can be implemented for the thin-film waveguide structures discussed in the article. The data have been obtained with the rigorously calculated $R_2(k_z)$ curves (the one in Fig. 2 and others similar to it) used as ''experimental'' curves. The quantity $\Delta k_z$ is chosen on the basis of the condition $R_2(k_z^{(0)}+\Delta k_z)=(1+R_2^{(0)})/2$. The first column of the table lists the conditional mode orders. The polarization of the modes is denoted by the symbols $\mathrm{TE}_p$ and $\mathrm{TM}_p$, where $p=W$ for waveguide modes and $p=L$ for leaky modes. We note that the quantity $|\Delta h|$ increases as $|\mathrm{Im}\, h|$ increases, lowering the accuracy of approximation of Eq. (18). The growth of $|\Delta h|$ follows from Eqs. (19)–(22) with allowance for the fact that high-contrast resonance troughs of the $R_2(k_z)$ curve can be obtained if $|p_2|\sim|\mathrm{Im}\, h|/k_0$. The latter estimate is readily obtained by comparing relations (19), (22), and (26) and taking the inequality $\sigma\ll 1$ into account. On the other hand, all the data in the table have been obtained for values of $g$ such that $R_2^{(0)}<0.75$. This remark accounts for the noticeable increase in the error of solution of the inverse problems as the losses of the selected modes increase. We also call attention to the possibility of reconstructing the parameters of single-mode (for a fixed polarization) waveguide films from the values of $h$ for waveguide and leaky modes (rows 4–6 in Table I). The accuracy of reconstruction in this case is higher than when $h$ is used for orthogonally polarized waveguide modes, because the system of dispersion relations is poorly conditioned in the latter case.

[1] S. H. Talisa, IEEE Trans. Microwave Theory Tech. **MTT-33**, 967 (1985).
[2] W. S. Borland, D. E. Zelman, C. Radens *et al.*, IEEE J. Quantum Electron. **QE-23**, 1172 (1987).
[3] C. Hulse and A. Knoeser, IEEE J. Quantum Electron. **QE-28**, 2682 (1992).
[4] J. K. Shaw, A. K. Jordan, and W. R. Winfrey, J. Opt. Soc. Am. A **10**, 1157 (1993).
[5] F. A. Burton and S. A. Cassidy, J. Lightwave Technol. **LT-8**, 1843 (1990).
[6] K. H. Schlereth and M. Tack, IEEE J. Quantum Electron. **QE-26**, 627 (1990).
[7] A. A. Romanenko, A. B. Sotskiĭ, and A. V. Khomchenko, IF ANB Preprint No. 649 [in Russian], (B. I. Stepanov Institute of Physics, Academy of Sciences of Belarus, Minsk (1991), 31 pp.
[8] M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable* [in Russian], Nauka, Moscow (1973), 736 pp.
[9] V. V. Shevchenko, Differents. Uravn. **15**, 2004 (1979).
[10] M. Adams, *An Introduction to Optical Waveguides* (Wiley, New York, 1982) [Russian trans., Mir, Moscow (1984), 512 pp.].
[11] R. Ulrich, J. Opt. Soc. Am. **60**, 1337 (1970).
[12] L. N. Deryugin, A. N. Marchuk, and V. E. Sotin, Izv. Vyssh. Uchebn. Zaved. Radioélektron. **13**, 973 (1970).
[13] B. P. Singh and P. N. Prasad, J. Opt. Soc. Am. B **5**, 453 (1988).
[14] A. K. Nikitin, A. A. Tishchenko, and A. I. Chernyaĭ, Zarub. Radioélektron., No. 10, 14 (1990).
[15] F. Yang and J. R. Sambles, J. Opt. Soc. Am. B **10**, 858 (1993).
[16] F. Yang, J. R. Sambles, and G. B. Bradberry, J. Appl. Phys. **78**, 2187 (1995).
[17] A. B. Sotskiĭ, A. V. Khomchenko, and L. I. Sotskaya, Opt. Spektrosk. **78**, 502 (1995) [Opt. Spectrosc. **78**, 453 (1995)].

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Method of eigenfunctions of singular operators in the theory of diffraction by a thick vibrator

A. V. Sochilin and S. I. Éminov

*Yaroslav Mudryĭ Novgorod State University, 173003 Novgorod, Russia*

(Submitted October 18, 1996)
Zh. Tekh. Fiz. **68**, 96–101 (April 1998)

A rigorous electrodynamic solution of the problem of the diffraction of electromagnetic waves by the surface of a vibrator is described by a system of integrodifferential equations. The method of eigenfunctions of singular operators is used to reduce the basic system to an infinite algebraic Fredholm system of the second kind. The high efficiency of the proposed method is demonstrated on concrete examples. © *1998 American Institute of Physics.* [S1063-7842(98)01704-8]

## INTRODUCTION

The solution of problems in the diffraction of electromagnetic waves by a thick vibrator and other arbitrary open surfaces reduces to the solution of systems of integrodifferential equations in the two tangential components of the density of surface currents. The solution of diffraction problems in this general setting is a difficult undertaking. So far numerical methods have been formulated, but there are no convergence theorems for them.

In the class of all arbitrary surfaces we distinguish surfaces formed by the rotation of a piecewise-smooth line: open surfaces of revolution. The most commonly encountered geometries in the theory of antennas are cylindrical, spherical, and conical surfaces of revolution.

It will be shown below that the problem for the indicated class of antennas can be reduced to a system of two one-dimensional integrodifferential equations, which admits a mathematically rigorous solution. The objective of the present study is to analyze the systems of integrodifferential equations.[1,2]

For brevity we confine the discussion to a cylindrical surface, but the results can be extended to any surface of revolution.

## 1. SINGULAR INTEGRAL AND INTEGRODIFFERENTIAL OPERATORS

We introduce the four singular operators

$$(Lu)(\tau)=\frac{1}{\pi}\int_{-1}^{1}\frac{u(t)}{\sqrt{1-t^2}}\ln\frac{1}{|\tau-t|}\,dt, \tag{1}$$

$$(SLu)(\tau)=\frac{1}{\pi}\frac{\partial}{\partial\tau}\int_{-1}^{1}\frac{u(t)}{\sqrt{1-t^2}}\ln\frac{1}{|\tau-t|}\,dt, \tag{2}$$

$$(SAu)(\tau)=\frac{1}{\pi}\int_{-1}^{1}u(t)\sqrt{1-t^2}\,\frac{\partial}{\partial t}\ln\frac{1}{|\tau-t|}\,dt, \tag{3}$$

$$(Au)(\tau)=\frac{1}{\pi}\frac{\partial}{\partial\tau}\int_{-1}^{1}u(t)\sqrt{1-t^2}\,\frac{\partial}{\partial t}\ln\frac{1}{|\tau-t|}\,dt. \tag{4}$$

The operators $L$ and $A$ have been investigated previously.[1,3] For the discussion that follows we need to select the spaces in which all four operators will act. We introduce a Hilbert weight space $L_{2,\rho_1}[-1,1]$ with the weighting function $\rho_1(\tau)=(1-\tau^2)^{-1/2}$ and the orthonormal basis

$$\varphi_n(\tau)=\begin{cases}\sqrt{1/\pi}, & n=1,\\ \sqrt{2/\pi}\,\cos[(n-1)\arccos(\tau)], & n=2,3,\ldots.\end{cases} \tag{5}$$

The operator $L$ acts in $L_{2,\rho_1}[-1,1]$, where the following relation holds:

$$(L_{\varphi_n})(\tau)=\lambda_n\varphi_n(\tau),\quad \lambda_1=\ln 2,$$
$$\lambda_n=\frac{1}{n-1},\quad n=2,3\ldots. \tag{6}$$

Together with $L_{2,\rho_1}[-1,1]$ we also introduce the weight space $L_{2,\rho_2}[-1,1]$ with the weighting function $\rho_2(\tau)=(1-\tau^2)^{1/2}$ and the orthonormal basis

$$\psi_n(\tau)=\frac{1}{\rho_2(\tau)}\sqrt{\frac{2}{\pi}}\,\sin[n\arccos(\tau)],\quad n=1,2,\ldots. \tag{7}$$

The operator $A$ acts in $L_{2,\rho_2}[-1,1]$ and, as shown in Ref. 1,

$$(A\psi_n)(\tau)=n\psi_n(\tau),\quad n=1,2,3,\ldots. \tag{8}$$

Next we consider the operator $SL$: $L_{2,\rho_1}\to L_{2,\rho_2}$. From the definition of this operator and relation (6) we have

$$(SL\varphi_1)(\tau)=0,\quad (SL\varphi_n)(\tau)=\psi_{n-1}(\tau),$$
$$n=2,3,\ldots. \tag{9}$$

Analogously, for the operator $SA$: $L_{2,\rho_2}\to L_{2,\rho_1}$ we obtain the following relation by first integrating by parts and then applying Eq. (6):

$$(SA\psi_n)(\tau)=\varphi_{n+1}(\tau),\quad \tau=1,2,\ldots. \tag{10}$$

We have thus completely described the eigenfunctions of the operators introduced above. We note incidentally that the

1063-7842/98/43(4)/5/$15.00

operator $L$ is completely continuous, the operators $SL$ and $SA$ are bounded, and finally the operator $A$ is symmetric, positive definite, and has a dense domain of definition. The operator $L$ arises in the problem of the diffraction of $E$-polarized waves by a perfectly conducting strip. The electric currents induced by a wave on both sides of the strip run parallel to the edge of the strip and increase without bound as the edge is approached. The function $\rho_1(\tau)$ describes this behavior.

The operator $A$ is encountered in the diffraction problem for $H$-polarized waves. In this case the currents run perpendicular to the edge and vanish as the edge is approach, as does the function $\rho_2(\tau)$. In the problem of diffraction by a cylindrical tubular vibrator the currents flow both parallel and perpendicular to the edge. And it will be evident below that all four singular operators arise.

## 2. REDUCTION OF A SYSTEM OF TWO-DIMENSIONAL INTEGRODIFFERENTIAL EQUATIONS TO ONE-DIMENSIONAL SYSTEMS

Under the influence of the primary field $(\mathbf{E}^0,\mathbf{H}^0)$ currents are induced on the surface of a tubular vibrator with the current density $\mathbf{j}(z,\varphi)=\mathbf{t}_z j_z(z,\varphi)+\mathbf{t}_\varphi j_\varphi(z,\varphi)$, where $\mathbf{t}_z$ and $\mathbf{t}_\varphi$ are unit vectors along the $z$ and $\varphi$ axes.

These currents generate a secondary field. The tangential component of the total electric field must vanish on the perfectly conducting surface of the vibrator. From this condition we deduce the system of integrodifferential equations

$$\int_S \left[ j_z \left( \frac{\partial^2 G}{\partial z \partial z'} - k^2 G \right) + j_\varphi \frac{\partial^2 G}{a \partial z \partial \varphi'} \right] dS' = i \sqrt{\frac{\varepsilon}{\mu}}\, E_z^0,$$

$$\int_S \left[ j_z \left( \frac{\partial^2 G}{a \partial z' \partial \varphi} \right) + j_\varphi \left( \frac{\partial^2 G}{a^2 \partial \varphi \partial \varphi'} - k^2 \cos(\varphi-\varphi') G \right) \right] dS'$$

$$= i \sqrt{\frac{\varepsilon}{\mu}}\, E_\varphi^0, \tag{11}$$

where

$$G = \frac{\exp(-ikR)}{4\pi kR} = \sum_{m=-\infty}^{+\infty} \exp(-im(\varphi-\varphi')) S_m(z,z'),$$

$$S_m = \frac{1}{2\pi^2} \int_0^{+\infty} \cos[kx(z-z')]$$

$$\times I_m(\sqrt{x^2-1}\,ka) k_m(\sqrt{x^2-1}\,ka)\, dx,$$

and $I_m$ and $K_m$ are the modified Bessel functions.

We now form Fourier-type expansions of the functions $j_z$, $j_\varphi$, $E_z^0$, and $E_\varphi^0$:

$$j_z = \frac{1}{2\pi a} \sum_{m=-\infty}^{+\infty} I_z^m(z) \exp(-im\varphi),$$

$$\int_0^{2\pi} j_z \exp(im\varphi') a\, d\varphi' = I_z^m(z'), \tag{12}$$

$$j_\varphi = \frac{1}{2\pi a} \sum_{m=-\infty}^{+\infty} I_\varphi^m(z) \exp(-im\varphi),$$

$$\int_0^{2\pi} j_\varphi \exp(im\varphi') a\, d\varphi' = I_\varphi^m(z'), \tag{13}$$

$$E_z^0 = \sum_{m=-\infty}^{+\infty} E_z^m(z) \exp(-im\varphi),$$

$$E_\varphi^0 = \sum_{m=-\infty}^{+\infty} E_\varphi^m(z) \exp(-im\varphi). \tag{14}$$

We also take into account the relations

$$\cos(\varphi-\varphi') = \frac{\exp(-i(\varphi-\varphi')) + \exp(i(\varphi-\varphi'))}{2}, \tag{15}$$

$$\cos(\varphi-\varphi') \sum_{m=-\infty}^{+\infty} S_m \exp(-im(\varphi-\varphi'))$$

$$= \frac{1}{2} \sum_{m=-\infty}^{+\infty} (S_{m-1}+S_{m+1}) \exp(-im(\varphi-\varphi')). \tag{16}$$

We use the expansions in Fourier series to reduce the system of two-dimensional integrodifferential equations (11) to one-dimensional systems. At the same time we transform to dimensionless variables. To simplify the writing, we assume that $l$ denotes the electrical length of the arm $kl$, and $a$ denotes the electrical radius of $ka$. As a result, we obtain the system

$$\int_{-1}^{1} \left[ I_z^m \left( \frac{1}{l} \frac{\partial^2 S_m}{\partial \tau \partial t} - l S_m \right) + I_\varphi^m \frac{im}{a} \frac{\partial S_m}{\partial \tau} \right] dt = \frac{i}{k} \sqrt{\frac{\varepsilon}{\mu}}\, E_z^m,$$

$$\int_{-1}^{1} \left[ I_z^m \frac{-im}{a} \frac{\partial S_m}{\partial t} + I_\varphi^m \left( \frac{m^2 l}{a^2} S_m - l \frac{S_{m-1}+S_{m+1}}{2} \right) \right] dt$$

$$= \frac{i}{k} \sqrt{\frac{\varepsilon}{\mu}}\, E_\varphi^m. \tag{17}$$

The system of one-dimensional integrodifferential equations (17) must be solved for values of $m$ for which at least one of the functions $E_z^m, E_\varphi^m$ is nonvanishing.

## 3. PARTITIONING THE PROBLEM INTO EVEN AND ODD PARTS; SEPARATION OF SINGULAR OPERATORS

In the representation of the function $S_m$ by the Fourier integral

$$S_m(\tau,t) = \frac{1}{2\pi^2} \int_0^{+\infty} \cos[lx(\tau-t)]$$

$$\times I_m(\sqrt{x^2-1}\,a) K_m(\sqrt{x^2-1}\,a)\, dx \tag{18}$$

the spectral density $I_m K_m$ is an even function in $x$. Making use of this fact along with the form of the system (17), we partition the latter into two systems. We write the functions $E_z^m$ and $E_\varphi^m$ in the form

$$E_z^m(z) = E_z^{m+}(z) + E_z^{m-}(z), \quad E_z^{m\pm}(\pm z) = \pm E_z^{m\pm}(z),$$

$$E_\varphi^m(z) = E_\varphi^{m+}(z) + E_\varphi^{m-}(z), \quad E_\varphi^{m\pm}(\pm z) = \pm E_\varphi^{m\pm}(z).$$

On the basis of this representation we expand the column formed by the right-hand sides of the system (17) into two columns:

$$\begin{pmatrix} E_z^m \\ E_\varphi^m \end{pmatrix} = \begin{pmatrix} E_z^{m+} \\ E_\varphi^{m-} \end{pmatrix} + \begin{pmatrix} E_z^{m-} \\ E_\varphi^{m+} \end{pmatrix}. \tag{19}$$

We use the representation (19) and the even property of the kernels to reduce the system (17) to two independent systems. In the first of these systems, which we call the even problem, the function $I_z^m$ is even, and the function $I_\varphi^m$ is odd. In the second system the situation is reversed. We consider the even problem below. The investigation of the odd problem is analogous.

Invoking the asymptotic forms of the modified Bessel functions

$$I_m(x)K_m(x) = \frac{1}{2x} + \frac{1}{16x^3} + \ldots, \quad x \to +\infty \tag{20}$$

we separate the logarithmic singularity of the function $S_m$:

$$S_m(\tau,t) = \frac{1}{4\pi^2 a} \ln \frac{1}{|\tau - t|} + N_m(\tau,t). \tag{21}$$

The properties of the function $N_m(\tau,t)$ follow from the representation (20). The first partial derivatives $\partial N_m/\partial\tau$ and $\partial N_m/\partial t$ are continuous functions. The second partial derivative $\partial^2 N_m/\partial\tau\partial t$ has a logarithmic singularity. Moreover, the expansion (20) contains only terms in odd powers of $(1/x)$. We transform from the functions $I_z^{m+}$ and $I_\varphi^{m-}$ to the new functions

$$I_z^{m+}(\tau) = \rho_2(\tau)u^+(\tau), \quad I_\varphi^{m-}(\tau) = \rho_1(\tau)\nu^-(\tau). \tag{22}$$

We omit the index $m$ below, because we are concerned with the system for fixed $m$. The functions $\rho_1(\tau)$ and $\rho_2(\tau)$ describe the behavior of the currents near the edge according to the Meixner condition. On the other hand, these functions guarantee orthogonality, as shown in Sec. 1.

We therefore trace the intimate relationship between the Meixner conditions at the edge and the eigenfunctions of the singular operators of problems in diffraction by thin screens. But now on the basis of the representation (21) we distinguish singular operators having a simple form for the four singular integrodifferential operators describing the system (17). We also write the system in the operator form

$$\begin{gathered} \frac{1}{l}Au^+ + N^{11}u^+ + \frac{im}{a}SL\nu^- + N^{12}\nu^- = e^+, \\ \frac{-im}{a}SAu^+ + N^{21}u^+ + \left(\frac{m^2 l}{a^2} - l\right)L\nu^- + N^{22}\nu^- = h^-. \end{gathered} \tag{23}$$

The operators $A$, $SL$, $SA$, and $L$ are defined in Sec. 1. The operators $N^{pq}$ $(p,q=1,2)$ are integral operators of the form

$$(N^{p1}u^+)(\tau) = \int_{-1}^{1} u^+(t)\rho_2(t)N^{p1}(\tau,t)dt, \quad p=1,2, \tag{24}$$

$$(N^{p2}\nu^-)(\tau) = \int_{-1}^{1} \nu^-(t)\rho_1(t)N^{p2}(\tau,t)dt, \quad p=1,2. \tag{25}$$

The properties of the function $N^{pq}(\tau,t)$ depend on the properties of the function $N_m(\tau,t)$. For example, the function $N^{11}(\tau,t)$ has a logarithmic singularity, $N^{12}(\tau,t)$ and $N^{21}(\tau,t)$ are continuous, and their partial derivatives have a logarithmic singularity. The properties of the function $N^{22}(\tau,t)$ are the same as those of $N_m(\tau,t)$.

We investigate the system (23) in the space $L_{2,\rho_2} \oplus L_{2,\rho_1}$, which is the direct sum of the Hilbert spaces $L_{2,\rho_2}$ and $L_{2,\rho_1}$.

We have thus obtained a system of one-dimensional integrodifferential equations and analyzed the structure of each singular operator separately.

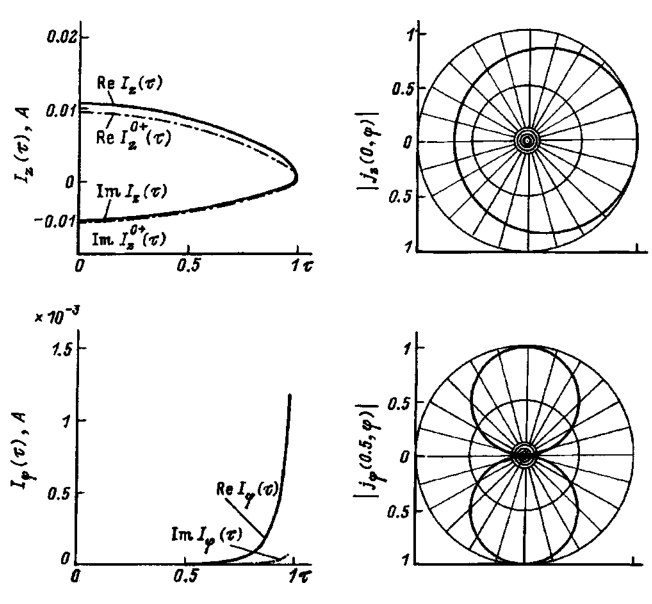

FIG. 1. $ka = \pi/20$, $2l = \lambda/2$.

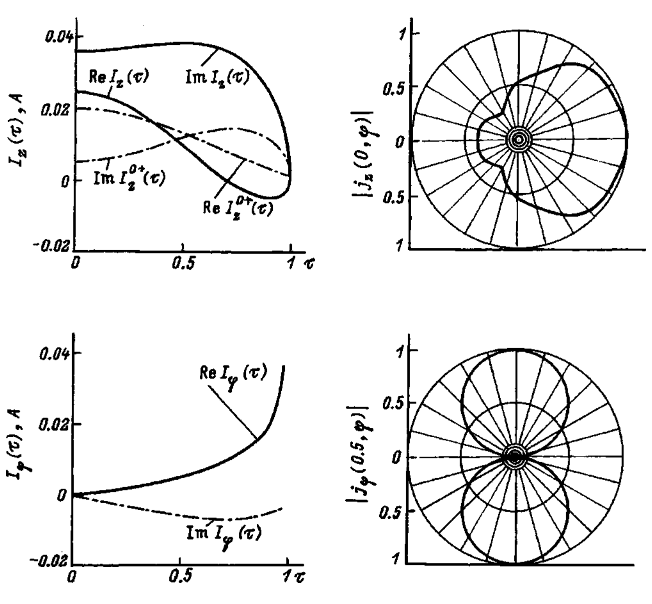

FIG. 2. $ka = \pi/2$, $2l = \lambda$.

TABLE I.

| | $ka=\pi/2,\ 2l=\lambda/2$ | | | |
|---|---|---|---|---|
| | $I_z^{1+}(\tau),A$ | | $I_\varphi^{1-}(\tau),A$ | |
| $N$ | Re $I_z^{1+}(\tau=0)$ | Im $I_z^{1+}(\tau=0)$ | Re $I_\varphi^{1-}(\tau=0.5)$ | Im $I_\varphi^{1-}(\tau=0.5)$ |
| 1 | $-0.566748e-02$ | $0.175142e-01$ | $0.522298e-02$ | $-0.602222e-02$ |
| 2 | $-0.535253e-02$ | $0.184216e-01$ | $0.579558e-02$ | $-0.722241e-02$ |
| 3 | $-0.535591e-02$ | $0.184216e-01$ | $0.578001e-02$ | $-0.722189e-02$ |
| 4 | $-0.535609e-02$ | $0.184550e-01$ | $0.578012e-02$ | $-0.722199e-02$ |
| 5 | $-0.535609e-02$ | $0.184550e-01$ | $0.578013e-02$ | $-0.722204e-02$ |
| 6 | $-0.535609e-02$ | $0.184550e-01$ | $0.578011e-02$ | $-0.722196e-02$ |
| 7 | $-0.535609e-02$ | $0.184550e-01$ | $0.578011e-02$ | $-0.722198e-02$ |
| 8 | $-0.535609e-02$ | $0.184550e-01$ | $0.578012e-02$ | $-0.722200e-02$ |
| 9 | $-0.535609e-02$ | $0.184550e-01$ | $0.578012e-02$ | $-0.722200e-02$ |
| 10 | $-0.535609e-02$ | $0.184550e-01$ | $0.578012e-02$ | $-0.722200e-02$ |
| | $ka=\pi/20,\ 2l=\lambda/2$ | | | |
| | $I_z^{1+}(\tau),A$ | | $I_\varphi^{1-}(\tau),A$ | |
| $N$ | Re $I_z^{1+}(\tau=0)$ | Im $I_z^{1+}(\tau=0)$ | Re $I_\varphi^{1-}(\tau=0.5)$ | Im $I_\varphi^{1-}(\tau=0.5)$ |
| 1 | $0.144909e-02$ | $0.268091e-04$ | $0.766654e-04$ | $-0.824154e-06$ |
| 2 | $0.118926e-02$ | $0.256874e-04$ | $-0.267393e-04$ | $-0.613851e-06$ |
| 3 | $0.126772e-02$ | $0.265931e-04$ | $0.472215e-05$ | $-0.234514e-06$ |
| 4 | $0.124767e-02$ | $0.263769e-04$ | $0.839380e-05$ | $-0.193556e-06$ |
| 5 | $0.125030e-02$ | $0.264077e-04$ | $0.575925e-05$ | $-0.224640e-06$ |
| 6 | $0.125100e-02$ | $0.264145e-04$ | $0.501021e-05$ | $-0.232928e-06$ |
| 7 | $0.125043e-02$ | $0.264086e-04$ | $0.514708e-05$ | $-0.231323e-06$ |
| 8 | $0.125061e-02$ | $0.264105e-04$ | $0.535972e-05$ | $-0.228936e-06$ |
| 9 | $0.125059e-02$ | $0.264103e-04$ | $0.531475e-05$ | $-0.229437e-06$ |
| 10 | $0.125058e-02$ | $0.264102e-04$ | $0.530634e-05$ | $-0.229540e-06$ |
| 11 | $0.125059e-02$ | $0.264103e-04$ | $0.530094e-05$ | $-0.229606e-06$ |

## 4. REDUCTION TO A FREDHOLM MATRIX SYSTEM OF THE SECOND KIND

To analyze the structure of the system (23), we transform from the operator system to an equivalent matrix system. We expand the unknown functions $u^{+}(\tau)$ and $\nu^{-}(\tau)$ in basis functions:

$$u^{+}(\tau)=\sum_{n=1}^{+\infty} c_n \frac{1}{\sqrt{2n-1}}\psi_{2n-1}(\tau), \tag{26}$$

$$\nu^{-}(\tau)=\sum_{n=1}^{+\infty} d_n \sqrt{2n-1}\,\varphi_{2n}(\tau). \tag{27}$$

Substituting Eqs. (26) and (27) into (23), we multiply the first equation of the system by the function $\psi_{2n-1}(\tau)/\sqrt{2n-1}$ in space $L_{2,\rho_2}$ and multiply the second equation by the function $\sqrt{2n-1}\,\varphi_{2n}(\tau)$ in space $L_{2,\rho_1}$. As a result, we obtain a matrix system to be analyzed in space $l_2 \oplus l_2$ ($l_2$ is the Hilbert space of sequences):

$$a^{11}c_n+\sum_{m=1}^{+\infty} c_m N_{mn}^{11}+a^{12}d_n+\sum_{m=1}^{+\infty} d_m N_{mn}^{12}=e_n,$$

$$a^{21}c_n+\sum_{m=1}^{+\infty} c_m N_{mn}^{21}+a^{22}d_n+\sum_{m=1}^{+\infty} d_m N_{mn}^{22}=h_n,$$

$$1 \leqslant n < +\infty, \tag{28}$$

where

$$a^{11}=\frac{1}{l}, \quad a^{12}=\frac{im}{a}, \quad a^{21}=-\frac{im}{a}, \quad a^{22}=\frac{m^2 l}{a^2}-l.$$

The basis functions are chosen in such a way as to make $a^{ij}$ independent of $n$. The numbers $N^{pq}$ $(p,q=1,2)$ are the matrix elements of the operators $N^{pq}$ in the corresponding Hilbert spaces, and the numbers $e_n$ and $h_n$ represent the coefficients of the expansion of the right-hand sides in the basis functions.

We note that for $m=0$, which corresponds to an axisymmetrically driven vibrator, the system (28) is separated into two independent matrix equations. For all other $m$ we transform the system (28). To do so, we solve the system for $c_n$ and $d_n$ and reduce it to the canonical form

$$c_n+\sum_{m=1}^{+\infty} c_m \hat{N}_{mn}^{11}+\sum_{m=1}^{+\infty} d_m \hat{N}_{mn}^{12}=\hat{e}_n,$$

$$d_n+\sum_{m=1}^{+\infty} c_m \hat{N}_{mn}^{21}+\sum_{m=1}^{+\infty} d_m \hat{N}_{mn}^{22}=\hat{h}_n. \tag{29}$$

It is noted at once that the matrix elements $\hat{N}_{mn}^{pq}$ are formed by linear combinations of the matrix elements $N_{mn}^{pq}$ and preserve the properties of these elements. The following theorem has been proved: The system of two matrix equations (29) is a Fredholm system of the second kind.

We conclude this section with the observation that the approach developed here is applicable to problems of diffraction by an arbitrary open surface of revolution.

## 5. RESULTS OF NUMERICAL CALCULATIONS

We consider the diffraction of a plane wave by a vibrator. The source of the primary field is a filament that carries an electric current of unit amplitude and is situated parallel to the axis of the vibrator at a large distance from it in comparison with the wavelength.

The efficiency of the method proposed in the article is demonstrated in Table I, which shows the results of calculations of the first-harmonic $(m=1)$ current. An analysis of the results shows that for $ka=\pi/2$ and $2l=\lambda/2$ ($\lambda$ is the wavelength) the value of $l_z^{1+}(\tau)$ converges with three-place accuracy for $N=2$ and with six-place accuracy for $N=4$ ($N$ is the number of basis functions). The value of $I_\varphi^{1-}(\tau)$ converges more slowly: to three and six significant figures for three and eight basis functions, respectively. An analysis of the convergence of the currents of other harmonics has also disclosed a high rate of convergence of the cutoff method.

Next, calculations have shown that only axial currents are induced on the surface of thin vibrators, and they are determined by the zeroth harmonic. Even for a vibrator of radius $a=\lambda/40$ azimuthal currents are faintly generated everywhere except at the ends of the vibrator and at the edges, where the currents become infinite in accordance with the Meixner condition (Fig. 1). A further increase in the vibrator produces the patterns shown in Fig. 2.

The following notation is used in the figures:

$$I_z(\tau)=\sum_{m=0}^{\infty}\varepsilon_m I_z^{m+}(\tau),\qquad \varepsilon_m=\begin{cases}1, & m=0,\\ 2, & m>0,\end{cases}$$

$$I_\varphi(\tau)=\sum_{m=1}^{\infty}2I_\varphi^{m-}(\tau).$$

The zeroth harmonic is no longer sufficient for determining the axial currents. The strongest excitation of axial currents takes place on the side where the primary wave is incident. The azimuthal current function increases simultaneously. The vibrator begins to radiate along its own axis, and this is the fundamental difference of nonaxisymmetric excitation of an electrically thick vibrator from axisymmetric excitation.

[1] V. N. Plotnikov, Yu. Yu. Radtsig, and S. I. Éminov, Zh. Vychisl. Mat. Mat. Fiz. **34**, 68 (1994).

[2] A. V. Sochilin and S. I. Eminov, in *Mathematical Methods in Electromagnetic Theory (MMET-94)* (Kharkov, 1994), pp. 426–429.

[3] Yu. Yu. Radtsig, A. V. Sochilin, and S. I. Éminov, Deposited at the All-Union Institute of Scientific and Technical Information, VINITI [in Russian], Deposit No. 2640-V93 (Moscow, 1993), 78 pp.

Translated by James S. Wood

# Influence of the beam-forming conditions on the development of space-charge oscillations in a long-pulse relativistic electron beam

L. Yu. Bogdanov, G. G. Sominskiĭ, and A. Ya. Fabirovskiĭ

*St. Petersburg State Technical University, 195251 St. Petersburg, Russia*

(Submitted October 21, 1996; resubmitted March 31, 1997)
Zh. Tekh. Fiz. **68**, 102–106 (April 1998)

The characteristics of the space-charge oscillations of a long-pulse relativistic electron beam in magnetically insulated diodes are determined for different geometries of the electron acceleration section and for explosive-emission cathodes of different materials. The important role of the stream of electrons having high transverse velocities in the evolution of the oscillations is demonstrated, and the laws governing the generation of this stream are determined. Possible mechanisms of the space-charge oscillations are described, taking into account the interaction of the electron stream in the beam halo with the main electron stream, the development of diocotron instability in the stream of electrons emitted by the outer lateral surface of the plasma emitter, and the instability of the space charge of ''long-lived'' electrons in the beam transport channel. © *1998 American Institute of Physics.* [S1063-7842(98)01804-2]

## INTRODUCTION

The vast and by no means exhausted potential applications of relativistic electron beams (REBs) in high-current electronics and power engineering, along with the search for effective techniques to control their characteristics, account for the ongoing fascination with this multifaceted research object. So far, unfortunately, it is often impossible to generate REBs of sufficiently high quality. One obstacle, in particular, is the shortage of information on the functional relations of the collective processes involved in such beams.

The laws of the formation and transport of REBS and their stability have been analyzed in a great many publications (see, e.g., Refs. 1–7). The main causes of the development of instabilities, primarily those associated with the existence of gradients of the velocities and density of electrons in the beams, have been discerned in theoretical studies. The mechanism underlying the onset of instability under the influence of these factors is fairly general, manifesting itself in phenomena other than REBs (see, e.g., Refs. 8 and 9). The published theoretical studies have been carried out for highly idealized beam models and largely in the linear approximation. They are successful in explaining the disintegration of REBs but are generally inapplicable for determining the amplitude-frequency and space-time response characteristics of the collective processes occurring in them.

In experimental work, owing to the difficulties of the diagnostics of high-current electron beams, the processes analyzed have been predominantly either very low-frequency ($\leqslant$100 MHz) or substantially higher-frequency ($\geqslant$3000 MHz) (see, e.g., Refs. 2, 10–13). Only recently through the application of a relatively sophisticated, relatively nondisruptive investigative technique has information been obtained on the characteristics of the collective processes over a broad frequency band.[14–18] The amplitude-frequency and space-time responses of oscillations have been measured in long-pulse REBs, which previously have been the least investigated kind. However, the data have not been adequate for unambiguously determining the nature of the beams. Here we report new results that lend better insight into the mechanisms of the collective processes and their relationship to the specific characteristics of the formation and transport of REBs in magnetically insulated diodes, which are widely used in practice.

## RESULTS OF MEASUREMENTS AND DISCUSSION

Measurements have been carried out using the SÉR-1 apparatus[15,19] in the experimental instrument shown in cross section in Fig. 1. Electrons emitted by the edge explosive-emission cathode *1* of diameter $D_c$ were accelerated in the gap $L_{ca}$ between the cathode and the anode *2* to an energy of approximately 200 keV, whereupon they drifted into the beam transport channel *3* (of diameter $D_t = 32$ mm and length $L_t = 1.2$ m) and were deposited on the surface of the collector *4* in the decaying edge magnetic field of the system of solenoids *5*. An approximately uniform magnetic field $B_0 \simeq 1$ T was created in the entire beam-transport space except in the edge zone. The deviations from $B_0$ at the solenoid junctions were less than 10%. The duration of the current pulses on the collector and the current intensity $I$ were varied in the ranges $\sim 1.0 - 3.0$ $\mu$s and $0.7 - 1.1$ kA, respectively.

Our primary concern in the study has been to ascertain the influence of the beam-forming conditions on the characteristics of the space-charge oscillations generated in the beam. Different cathode materials were used (stainless steel and carbon), along with different cathode diameters (12 mm and 20 mm), and the distance $L_{ca}$ between the cathode and the anode was varied from 12 mm to 27 mm. Moreover, the influence of local magnetic-field nonuniformities created by the auxiliary coils *6* and *7* situated at distances $\sim$20 cm and $\sim$60 cm from the cathode was tested in all the investigated systems. The coils produced magnetic-field nonuniformities having a corkscrew configuration with a half-width along the

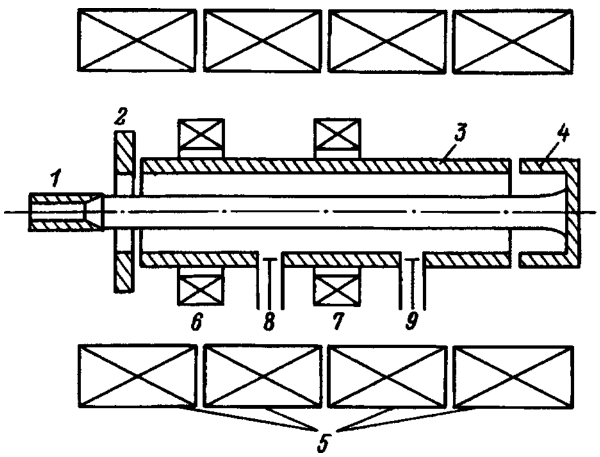

FIG. 1. Schematic view of the cross section of the experimental instrument.

axis of the system ~10 cm, and their amplitude in different experiments was equal to 30% or 60% of the main beam-confining field $B_0$.

The characteristics of the oscillations were determined by the procedure developed in Ref. 15 with the use of the probes *8* and *9*, which were set up at distances from the cathode of ~40 cm and ~80 cm, respectively. The probes ''communicated'' with the beam through small (diameter ~3 mm) openings in the wall of the transport channel and detected the induced signals associated with the motion of the space-charge packets in the beam. The magnetic spectra of the oscillations in a period ~6–15 ns and the pulsed variations of the detected probe pulses were recorded.

The characteristics of the oscillations are qualitatively similar in all the investigated systems. The variations of the oscillation amplitude $A$ with the time $t$ and in space can be traced by observing the detected probe signals. Figure 2 shows typical $A(t)$ curves of the probes 8 and 9. Also shown are the time variations of the collector current $I$. Our previous investigations[15] have shown that the probes are coupled with the near field of the transport channel and reliably detect REB space-charge oscillations until the dense cathode plasma, undergoing axial expansion, reaches the level of the entrance openings into the probe section.

The plasma oscillations are scarcely detected at all by the probe *9* during the entire duration of the beam current pulse, but they are what determine the signal of the probe *8* in time intervals corresponding to the decay of the current on the collector. At the leading edge of the current $I$ pulse we discern time segments characterized by buildup of the space-charge oscillations with increasing distance from the cathode. Their amplitude at the probe *9* attains values $\sim 10^6$ V/m in all the investigated systems (i.e., ~10% of the static field between the beam and the wall of the transport channel). These data indicate amplification of the wave processes in the interval between the probes and also that the output signals of the probe far from the cathode are determined, in the given time interval at any rate, by growing waves in the beam. A lack of data on the spatial variation of the oscillation amplitude in the near field of the cathode makes it difficult to assess the nature of the oscillations of the probe *8*. In principle, its signals can be attributed not only to the amplification of waves propagating along the beam, but also to oscillatory processes of another kind. Consequently, to determine the relations characterizing the growth of the waves in the beam, we analyze the prominent features of the signal characteristics of the probe far from the cathode.

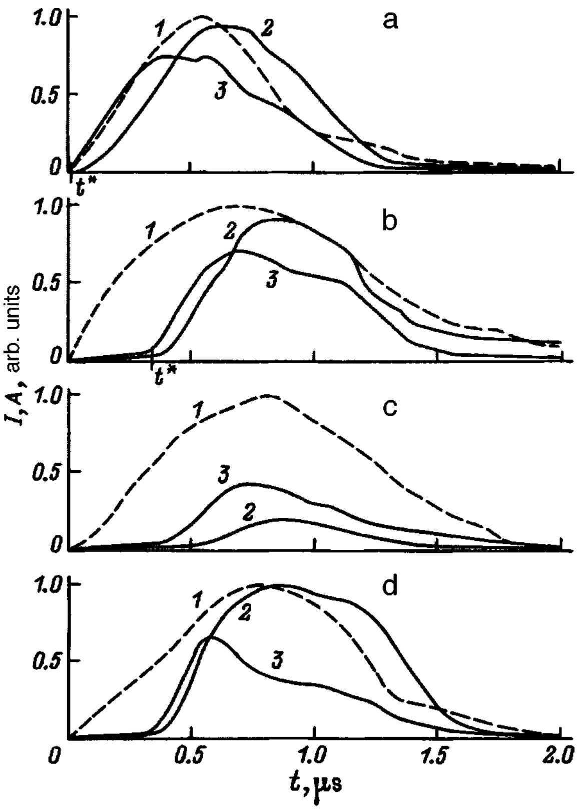

FIG. 2. Time variations of: *1*) the collector current $I$; *2,3*) the amplitude $A$ of the detected signals from the probes *8* and *9*, respectively, in a relativistic electron beam-forming system with a graphite cathode; a,b) characteristic curves obtained with confinement of the REB in a uniform magnetic field $B_0 \cong 1$ T; c,d) data obtained with the introduction of local magnetic-field nonuniformities of amplitude $0.3B_0$ by means of the coils *6* and *7*, respectively; a) $D_c = 20$ mm, $L_{ca} = 12$ mm; b,c,d($D_c = 20$ mm, $L_{ca} = 27$ mm.

As the beam current is increased, the signal amplitudes of the probe *9* increase with time, initially at a slow rate, but then after a certain time interval $t = t^*$ specific to the given system the rate of the process increases abruptly (Fig. 2). The quantity $t^*$ changes considerably when the geometry of the electron acceleration section or the material of the cathode is altered. Table I shows typical values of $t^*$ for the relativistic electron beam-forming systems investigated here.

TABLE I.

| | Cathode material | | | | |
|---|---|---|---|---|---|
| Characteristic | Stainless steel | Carbon | | | |
| $D_c$, mm | 20 | 12 | | 20 | |
| $L_{ca}$, mm | 27 | 12 | 27 | 12 | 27 |
| $t^*$, $\mu$s | 0.5–0.6 | 0.2 | 0.4–0.5 | 0–0.1 | 0.3–0.4 |
| $t_d$, $\mu$s | 0.6 | 0.15 | 0.4 | 0.1 | 0.35 |

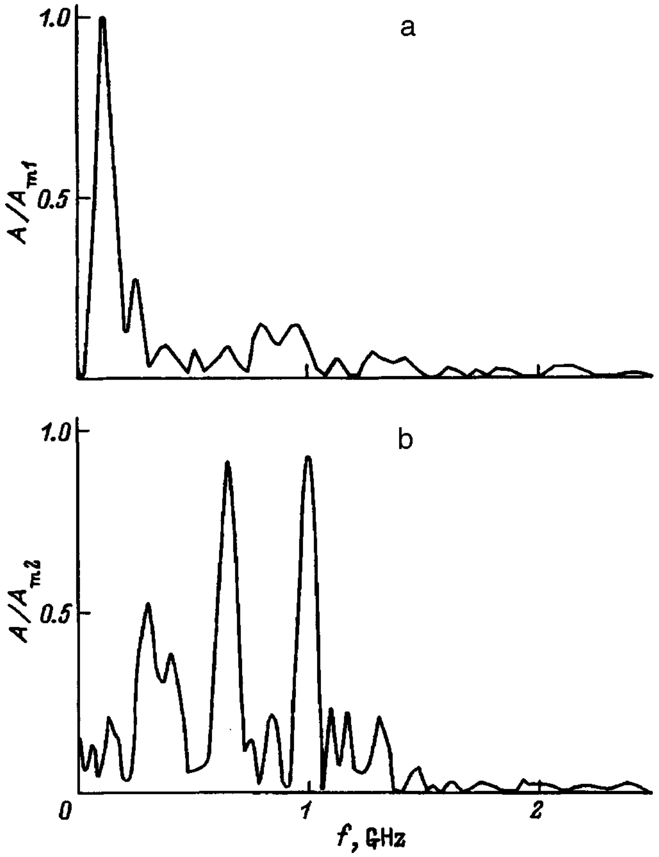

FIG. 3. Typical spectra of signals from the probe *9*, obtained in a uniform magnetic field $B_0 \cong 1$ T in a relativistic electron beam-forming system with a graphite cathode, $D_c = 20$ mm, $L_{ca} = 27$ mm. a) $t = 0.3$ $\mu$s ($t < t^*$); b) $t = 0.6$ $\mu$s ($t > t^*$); $A_{m1}$ and $A_{m2}$ are the maximum values of the amplitude $A$ of the recorded spectra, $A_{m2}/A_{m1} \cong 15$.

The spectrum of the oscillations of the probe *9* also changes with time. At the initial times $t < t^*$, when the signal amplitudes are small, comparatively low-frequency oscillations in the frequency band 100 MHz $\leqslant f \leqslant$ 300 MHz are prevalent. But then in the presence of larger amplitudes (at $t > t^*$) discrete peaks are discernible in the spectrum in the range of higher frequencies approximately from 500 MHz to 1500 MHz. Typical spectra of the oscillations of the probe *9* are shown in Fig. 3.

The influence of local magnetic-field nonuniformities having a corkscrew configuration on the REB space-charge oscillations has been observed previously.[16–18] The measurements were performed in a relativistic electron beam-forming system with a stainless steel cathode of diameter $D_c = 20$ mm mounted at a distance $L_{ca} = 27$ mm from the entrance to the transport channel. The data obtained in the present study can be used to generalize the conclusions of the first measurements to different geometries of the electron acceleration section and to different cathode materials. The effect depends not only on the magnitude of the nonuniformity, but also on its position in the transport channel. A 30% increase in the magnetic field near the cathode (by means of the coil *6*) suppresses the signals of both probes. The signals of the probe *8* decrease considerably in time intervals when they are capable of coupling with the cathode plasma. In the vicinity of the probe *9*, on the other hand, there is a noticeable drop in the intensity of the REB space-charge oscillations. Here the oscillation amplitude decreases by as much as 1.5–4 times (Fig. 2c). Increasing this nonuniformity of the magnetic field to 60% of $B_0$ produces an additional two- to fourfold reduction of the oscillation amplitude of the probe *9*. The introduction of such nonuniformities by means of the coil *7* at a large distance from the cathode, in contrast with the above-described influence of the coil *6*, essentially does nothing more than accelerate the development of the oscillations, not only in the beam propagating past it, but also at the probe *8* (Fig. 2d).

TABLE II.

| Cathode material | $V_{p\perp}$, cm/s | $V_{p\parallel}$, cm/s |
|---|---|---|
| Stainless steel | $3\times10^5$ | $3\times10^6$ |
| Carbon | $5\times10^5$ | $5\times10^6$ |

The observed influence of magnetic-field nonuniformities can be explained if we assume that their generation causes some of the electrons in the beam to be reflected, which has a strong influence on the characteristics of the recorded oscillations. Taking into account the adiabatic variation of the magnetic field near the auxiliary coils *6* and *7*, we can use the expression for the adiabatic invariant to determine the reflection conditions, writing

$$V_{\perp 0}/V_{\parallel 0} \geqslant (B_0/\Delta B)^{1/2}, \tag{1}$$

where $V_{\perp 0}$ and $V_{\parallel 0}$ are the transverse and longitudinal components of the velocity of electrons striking the magnetic mirror, and $\Delta B$ is the amplitude of the magnetic field.

Since the influence of the magnetic-field nonuniformities begins to be felt at $B_0/\Delta B = 3$, it follows from Eq. (1) that electrons having transverse velocities $V_{\perp 0} \geqslant 1.7 V_{\parallel 0}$ must be reflected. Such velocities can be acquired by electrons exiting from the cathode plasma during motion in an electric field transverse to the magnetic field $E_\perp \sim 10^8$ V/m. To determine the conditions for the generation of electrons with such high transverse velocities, we have estimated the fields at the surface of a plasma emitter having various dimensions and axial positions, taking into account the irregularities of its surface.[1)] The estimates show that fields $E_\perp$ of the necessary strength can occur at irregularities of the outer lateral boundary of the plasma near the entrance to the transport channel, i.e., with a delay relative to the beginning of the current pulse, the length of the delay depending on the distance from this zone and the rate of expansion of the cathode plasma. The resulting data have enabled us to calculate the delay time $t_d$ for the investigated systems (Table I). It has been assumed in the calculations that the gain of the field amplified at irregularities of the plasma emitter is identical and equal to 5 for all the systems. The values used for the transverse ($V_{p\perp}$) and longitudinal ($V_{p\parallel}$) plasma velocities for stainless steel and carbon cathodes are given in Table II.

They have been selected in accordance with published data on the motion of a cathode plasma in the vicinity of the cathode.[7]

A comparison of the calculated values of $t_d$ with the experimentally measured delay time $t^*$ to abrupt growth of the probe *9* signals shows that these times are close for all the investigated systems. It seems clear to us that this agreement confirms the important role of electrons with high transverse velocities in the development of convective instability, which, as mentioned, determines the characteristics $A(t)$ in the region of the transport channel far from the cathode. Such electrons, judging from the above estimates, exist only in the halo of the REB. Zaĭtsev *et al.*[20] have previously observed the influence of the electron stream in the REB halo on the expansion of the wall of an annular beam in a magnetic field. The authors of the paper postulated that the observed phenomenon is associated with the development of instability, but did not specify its mechanism. Our data can be used to determine the most probable attributes of this mechanism.

The current of electrons with high transverse velocities is weak in comparison with the beam current. It would be difficult in this light to assume that the experimentally observed strong observations are associated with processes strictly in the periphery of the beam. A possible cause of the generation of oscillations, it seems to us, is the development of a two-stream instability[8,9] associated with the interaction of electrons in the periphery of the beam having the highest transverse velocities (and, accordingly the lowest longitudinal velocities $V_{\parallel\min}$) with electrons coming from the end of the plasma emitter and having the maximum longitudinal velocities $V_{\parallel\max}$. This kind of interaction should be most efficient when a fast space-charge wave in the first stream is in synchronism with a slow space-charge wave in the second stream. In this case, equating the phase velocities of the waves, we readily obtain an expression for the frequency $f$ of the resulting development of instability of the oscillations:

$$f=\frac{1}{2\pi}\times\frac{V_{\parallel\min}\omega_{p\ \max}+V_{\parallel\max}\omega_{p\ \min}}{V_{\parallel\max}-V_{\parallel\min}}, \tag{2}$$

where $\omega_{p\ \max}$ and $\omega_{p\ \min}$ are the reduced plasma frequencies of the streams with velocities $V_{\parallel\max}$) and $V_{\parallel\min}$, respectively.

Calculations using our previous data on the characteristics of the space-charge distribution in REBs[16–18] to determine the reduced plasma frequencies give values $f\sim500-700$ MHz, which roughly correspond to the strongest oscillations observed in the experiments.

The investigated two-stream model explains not only the range of frequencies of the observed oscillations, but also the significant difference in how their characteristics are affected by magnetic-field nonuniformities generated in different parts of the transport channel. The amplification of the oscillations upon activation of the distant (from the cathode) coil *7* can be attributed, for example, to the fact that the interaction of the stream of electrons emanating from the cathode with electrons reflected by the magnetic mirror begins to exert an appreciable influence in this case. Similarly, the interaction associated with activation of the coil *6* can be rendered ineffective by the substantially shorter length of the section between the cathode and the near coil.

The explanation of the space-charge oscillations observed in our work as the result of the above-described two-stream interaction mechanism is not the only one possible. Of the alternatives, we lean toward a mechanism associated with the development of diocotron instability in the stream of electrons emitted from the side of the plasma emitter. In this case the REB halo must acquire space-charge inhomogeneities rotating in the azimuth direction. The frequencies $f_m$ of these magnetron-type oscillations should increase when the electron drift velocity $V_d=E_\perp/B_0$ above the surface of the plasma emitter and the mode order of the oscillations $n_m=1, 2, 3, \ldots$ $(f_m\sim n_m\times V_d)$ increase. In the transport channel with a potential difference $\Delta U=200$ kV, $B_0=1$ T, and $n_m=1$ these waves are characterized by a frequency $f_m\sim600$ MHz.

Azimuthal oscillations are also possible in the space charge of electrons arrested in the special kind of trap formed between the cathode and the magnetic mirror. In view of the potential drop $\Delta U\sim50$ kV between the beam and the wall of the transport channel, the frequencies of the azimuthal space-charge oscillations in this case should have a value $f_t\sim n_t\cdot150$ MHz, where $n_t=1, 2, 3, \ldots$ is the mode order of the oscillations in the trap. Oscillations of this kind can evolve, in principle, when the magnetic field is devoid of nonuniformities responsible for electron reflections. As mentioned previously, for example, in Ref. 21, it is also possible for space charge to build up and be trapped for a long time in the REB transport channel as a result of some of the electrons being reflected first from the collector and then again by the electric field at the cathode.

The relative role of the modes of oscillation discussed above is difficult to assess on the basis of existing experimental data and estimates. A realistic scenario appears to be one in which the collective processes in REBs are three-dimensional in character, different modes of oscillatory motion coexist, and a small change in the characteristics of the beam or the beam transport channel is accompanied by energy transfer from one degree of freedom of oscillatory motion to another in a process similar to that observed previously in magnetron-type systems with the axial motion of electrons taken into account.[22–24]

## CONCLUSION

Following is a summary of the most important results of the study.

We have determined the amplitude-frequency response characteristics of the space-charge oscillations of relativistic electron beams in magnetically insulated diodes for different geometries of the electron acceleration section and for explosive-emission cathodes of different materials.

We have demonstrated the important role of the stream of electrons having high transverse velocities in the evolution of oscillations. We have also determined the functional relations involved in the formation of this stream.

We have described possible mechanisms of the space-charge oscillations, taking into account the interaction of the electron stream in the halo of the beam with the main electron stream, the development of diocotron instability in the stream of electrons emitted by the outer lateral surface of the plasma emitter, and the instability of the space charge of ''long-lived'' electrons in the REB transport channel.

The authors are grateful to V. A. Solntsev and A. N. Rodionov of the Moscow Institute of Electronic Machinery Construction (MIÉM) for furnishing the results of calculations used in analyzing the electrodes at the cathode.

This work has received support from the Russian Fund for Fundamental Research.

[1)] The results of numerical calculations of the field distribution in the vicinity of the plasma emitter by V. A. Solntsev and A. N. Rodionov of the Moscow Institute of Electronic Machinery Construction (MIÉM), together with published information[2] on the irregularities of the emitter surface, have been used in the calculations. The calculations have been carried out without regard for the space-charge fields of the emitter electrons and on the assumption of infinite conductivity of the plasma emitter.

---

[1] A. A. Rukhadze, L. S. Bogdankevich, S. E. Rosinskiĭ, and V. G. Rukhlin, *Physics of High-Current Relativistic Electron Beams* [in Russian], Atomizdat, Moscow, 1980, 165 pp.

[2] S. P. Bugaev, V. I. Kanavets, V. I. Koshelev, and V. A. Cherepenin, *Relativistic Multimode Microwave Oscillators* [in Russian], Nauka, Novosibirsk, 1991, 296 pp.

[3] V. S. Ivanov, S. I. Krementsov, M. D. Raĭzer *et al.*, Fiz. Plazmy **7**, 784 (1981) [Sov. J. Plasma Phys. **7**, 430 (1981)].

[4] V. E. Nechaev, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **25**, 1067 (1982).

[5] O. B. Ovsyannikova, F. F. Kamanets, and V. G. Leĭman, Radiotekh. Elektron. **29**, 1985 (1984).

[6] A. I. Eremeev, V. S. Kabanov, P. S. Strelkov *et al.*, Fiz. Plazmy **14**, 958 (1988) [Sov. J. Plasma Phys. **14**, 562 (1988)].

[7] N. E. Rozanov, Zh. Tekh. Fiz. **61**(11), 151 (1991) [Sov. Phys. Tech. Phys. **36**, 1281 (1991)].

[8] V. N. Shevchuk, G. N. Shvedov, and A. V. Soboleva, *Wave and Oscillatory Phenomena in Microwave Electron Streams* [in Russian], Izd. SGU, Saratov, 1962, 335 pp.

[9] M. I. Rabinovich and D. I. Trubetskov, *Introduction to the Theory of Oscillations and Waves* [in Russian], Nauka, Moscow, 1984, 432 pp.

[10] A. F. Aleksandrov, S. Yu. Galuzo, V. I. Kanavets *et al.*, Zh. Tekh. Fiz. **52**, 110 (1982) [Sov. Phys. Tech. Phys. **27**, 71 (1982)].

[11] S. P. Bugaev, M. P. Deĭchuli, V. I. Kanavets *et al.*, Radiotekh. Elektron. **29**, 55 (1984).

[12] A. G. Zherlitsin, G. V. Mel'nikov, and G. P. Fomenko, Zh. Tekh. Fiz. **57**, 546 (1987) [Sov. Phys. Tech. Phys. **32**, 328 (1987)].

[13] S. S. Pushkarev, V. A. Bondarenko, N. P. Gadetskiĭ *et al.*, Prib. Tekh. Eksp. No. **9**, 31 (1989).

[14] L. Yu. Bogdanov, S. V. Voskresenskiĭ, and G. G. Sominskiĭ, in *Problems in Physical Electronics* [in Russian] (Izd. LIYaF, Leningrad, 1991), pp. 102–119.

[15] L. Yu. Bogdanov and G. G. Sominskiĭ, Zh. Tekh. Fiz. **65**(12), 77 (1995) [Tech. Phys. **40**, 1245 (1995)].

[16] A. V. Arkhipov, L. Yu. Bogdanov, S. V. Voskresenskiĭ *et al.*, in *Lectures on Microwave Electronics and Radiophysics (Tenth Winter School-Seminar of Engineers)* [in Russian], Vol. 2 (Izd. GosUNTs ''Kolledzh,'' 1996), pp. 3–34.

[17] L. Yu. Bogdanov and G. G. Sominski, in *Proceedings of the 11th International Conference on High Power Particle Beams (BEAMS'96)*, Vol. 1 (Prague, 1996), pp. 319–322.

[18] L. Yu. Bogdanov, G. G. Sominskiĭ, and S. V. Voskresenskiĭ, Zh. Tekh. Fiz. **67**(8), 84 (1997) [Tech. Phys. **42**, 930 (1997)].

[19] A. V. Arkhipov, L. Yu. Bogdanov, S. V. Voskresenkiĭ *et al.*, in *Problems in Physical Electronics* [in Russian] (Izd. LIYaF, Leningrad, 1989), pp. 28–47.

[20] N. I. Zaĭtsev, E. V. Ilyakov, G. S. Korablev *et al.*, Zh. Tekh. Fiz. **61**(12), 100 (1991) [Sov. Phys. Tech. Phys. **36**, 1387 (1991)].

[21] N. I. Zaĭtsev, G. S. Korablev, I. S. Kulagin, and V. E. Nechaev, Fiz. Plazmy **8**, 918 (1982) [Sov. J. Plasma Phys. **8**, 515 (1982)].

[22] G. G. Sominskiĭ, Pis'ma Zh. Tekh. Fiz. **3**, 181 (1977) [Sov. Tech. Phys. Lett. **3**, 72 (1977)].

[23] G. G. Sominskiĭ, *Lectures on Microwave Electronics (Fourth Winter School-Seminar of Engineers)* [in Russian], Vol. 3 (Izd. SGU, Saratov, 1978), pp. 119–138.

[24] S. A. Levchuk, G. G. Sominskiĭ, and S. V. Voskresenskiĭ, Pis'ma Zh. Tekh. Fiz. **14**, 1194 (1988) [Sov. Tech. Phys. Lett. **14**, 523 (1988)].

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Calculation of the focusing and aberrations of charged-particle beams in polar toroidal analyzers

M. I. Yavor

*Institute of Analytical Instrumentation, Russian Academy of Sciences, 198103 St. Petersburg, Russia*

(Submitted November 6, 1996)
Zh. Tekh. Fiz. **68**, 107–111 (April 1998)

A method is proposed for investigating the spatial and time-of-flight properties of polar toroidal charged-particle analyzers on the basis of calculations of the aberration integrals. The efficiency of the method is illustrated by comparing the results of calculations based on it with the results of numerical simulation and experimental measurements. © *1998 American Institute of Physics.* [S1063-7842(98)01904-7]

Polar toroidal electrostatic charged-particle analyzers (toroidal capacitors in which the particle beam moves in the meridional direction) are widely used in systems for the static and time-of-flight analysis of electron or ion beams with respect to angle and energy simultaneously[1–3] or with respect to mass.[4,5] In the past, new charged-particle beam analyzers containing polar toroidal capacitors either exclusively or in combination with other elements[6,7] have been designed by the time-consuming numerical simulation of charged-particle paths[8–10] or by means of crude, semianalytical approximations,[4,11] making it difficult to obtain optimal solutions requiring the simultaneous satisfaction of many focusing conditions in the presence of stringent geometrical constraints. The analytical method developed in Ref. 12 for calculating the electrostatic field in a polar toroidal capacitor suggests an effective approach to the investigation of the spatial and time-of-flight properties of such a capacitor, based on the computation of aberration integrals.

In this paper we investigate particle motion in the field of a capacitor without regard for edge effects at its input and output. These effects can be calculated on the basis of a method proposed earlier.[13]

We introduce toroidal coordinates $[r,\Phi,\Theta]$ (Fig. 1), where $\Theta$ is the angle of rotation about the symmetry axis ($Z$ axis) of the system, and $r$ and $\Phi$ are polar coordinates in the meridian plane relative to a point situated at a distance $a$ from the $Z$ axis (this point is assumed to be the center of curvature of the electrodes of the polar toroidal capacitor in the meridian plane). The surfaces of the electrodes of the polar toroidal capacitor are described by the equations $r_{1,2}=r_0\mp b$, where $2b$ is the interelectrode gap. An analytical expression for the electrostatic potential $\Psi(r,\Phi)$ between the capacitor electrodes, across which potentials of equal value and opposite sign to $V$ are specified, has been obtained[12] in the form of the expansion

$$\Psi(u,\Phi)=V\sum_{j=1}^{\infty}\Psi_j(u,\Phi)\varepsilon^j \tag{1}$$

in powers of the parameter $\varepsilon=b/r_0$. For real analyzer configurations this parameters is small, so that the potential of the field is very accurately described by the first few terms of the expansion (1), whose coefficients are equal to

$$\Psi_0(u,\Phi)=u,\quad \Psi_1(u,\Phi)=-\frac{1}{2}(1+C)(u^2-1),$$

$$\Psi_2(u,\Phi)=\frac{1}{3}(1+C+C^2)(u^3-u),$$

$$\Psi_3(u,\Phi)=\left[-\frac{1}{4}(1+C)(1+C^2)-\frac{1}{24}(1-C)(C+S^2)\right]\times(u^4-1)+\frac{1}{2}\left[\frac{1}{3}(1+C)(1+C+C^2)+\frac{1}{2}(1-C)(C+S^2)\right](u^2-1),$$

$$\Psi_4(u,\Phi)=\frac{1}{60}\left[12+\frac{33}{2}C+9C^2+\frac{19}{2}C^3+12C^4+4S^2-3C^2S^2+CS^2\right](u^5-u)+\frac{1}{18}\left[-2-\frac{19}{2}C-3C^2-\frac{1}{2}C^3-2C^4-5S^2+4S^2C^2+CS^2\right]\times(u^3-u).$$

Here we have introduced the notation $C=\cos\Phi/F$, $S=\sin\Phi/F$, $F=d+\cos\Phi$, $d=a/r_0$. We write the equations for the charged-particle paths in the polar toroidal capacitor in cylindrical coordinates, because the initial conditions for these paths are easier to determine in linear rather than angular variables. We introduce cylindrical coordinates $[\rho,\varphi,\bar{y}]$ with the axis of the coordinate system oriented perpendicular to the meridian plane $\Theta=0$ and passing through the center of polar curvature of the electrode surfaces in this plane (i.e., situated at a distance $a$ from the symmetry axis of the system). In the indicated meridian plane we have $\bar{y}=0$, $\rho=r$, and $\varphi=\Phi$, and the relations between the cylindrical and toroidal coordinates are described by the equations

$$r=\sqrt{\rho^2\sin^2\varphi+[\sqrt{\bar{y}^2+(a+\rho\cos\varphi)^2}-a]^2},$$

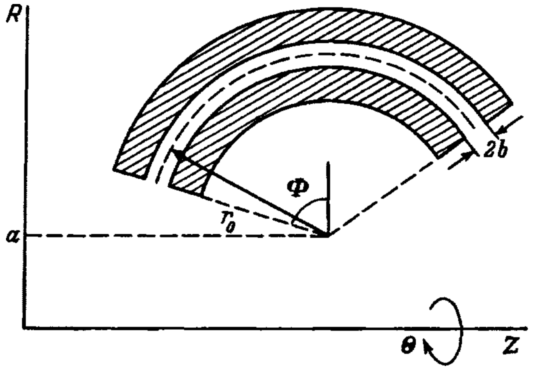

FIG. 1. Cross section of the electrodes of a polar toroidal capacitor in the meridian plane. Only the upper part of the cross section is shown, as it is symmetrical about the $Z$ axis.

$$\tan\Phi=\frac{\rho\sin\varphi}{\sqrt{\bar{y}^2+(a+\rho\cos\varphi)^2}-a},$$

$$\tan\Theta=\frac{\bar{y}}{a+\rho\cos\varphi}.$$

We denote $x=(\rho-r_0)/r_0$, $y=\bar{y}/r_0$ and expand the expressions for the toroidal coordinates $u$ and $\Phi$ in powers of $x$ and $y$. To within third-order terms these expansions have the form

$$\varepsilon u=x+\frac{1}{2}\,u_{yy}y^2+\frac{1}{2}\,u_{xyy}xy^2,$$

$$\Phi=\varphi+\frac{1}{2}\,\Phi_{yy}y^2+\frac{1}{2}\,\Phi_{xyy}xy^2, \tag{2}$$

where $u_{yy}=\cos\varphi/f$, $u_{xyy}=-\cos^2\varphi/f^2$, $\Phi_{yy}=-\sin\varphi/f$, $\Phi_{xyy}=\sin\varphi(d+2\cos\varphi)/f^2$, and $f=d+\cos\varphi$.

We now introduce the concept of a "main" particle as one that moves around a circular path of radius $r_0$ in the capacitor in the limit $\varepsilon\to 0$. Let $K_0$ be the kinetic energy of this particle (which is related to the potential $V$ by the equation $V=2\varepsilon K_0/q$), let $m_0$ be its mass, and let $q$ be the charge; we refer to these quantities as the nominal values of the corresponding quantities. The exact equations for the charged-particle paths in the polar toroidal capacitor are then written in the form

$$x''-\frac{2x'^2}{1+x}-1-x=\left[-\frac{\partial\Psi}{\partial x}+\frac{x'}{(1+x)^2}\,\frac{\partial\Psi}{\partial\varphi}\right]\times\frac{(1+x)^2+x'^2+y'^2}{1+\delta-2\Psi}, \tag{3}$$

$$y''-\frac{2x'y'}{1+x}=\left[-\frac{\partial\Psi}{\partial y}+\frac{y'}{(1+x)^2}\,\frac{\partial\Psi}{\partial\varphi}\right]\frac{(1+x)^2+x'^2+y'^2}{1+\delta-2\Psi}, \tag{4}$$

where $\delta$ is the deviation of the energy of an arbitrary particle from its nominal value, and the prime signifies differentiation with respect to the angle $\varphi$.

The equation for the normalized time of flight $\tau=\sqrt{2K_0/(m_0r_0^2)}\,(r-t_0)$, where $t$ is the time for an arbitrary particle to traverse the sector described by the angle of rotation $\varphi$, and $t_0=\varphi\sqrt{m_0r_0^2/2K_0}$ is the nominal time of flight of a main particle over an arc of radius $r_0$ of the same sector, has the form

$$\tau'+1=\sqrt{\frac{(1+x)^2+x'^2+y'^2}{1+\delta-2\Psi}(1+\gamma)}, \tag{5}$$

where $\gamma$ is the relative deviation of the mass of an arbitrary particle from its nominal value.

Substituting the equation (1) for the potential distribution, rewritten in the cylindrical coordinates (2), into the path and time-of-flight equations (3)–(5), we arrive at the following equations for the terms of the first $[x_1,y_1,\tau_1]$ and second $[x_2,y_2,\tau_2]$ aberration orders describing the coordinates of the charged-particle paths and the time of flight in the polar toroidal capacitor:

$$x_1''-a_{x'}x_1'-a_xx_1=a_\delta\delta+a_0, \tag{6}$$

$$y_1''-a_{y'}y_1'-a_yy_1=0, \tag{7}$$

$$\tau_1'=c_xx_1+c_\delta\delta+c_\gamma\gamma+c_0, \tag{8}$$

$$x_2''-a_{x'}x_2'-a_xx_2=a_{x'x'}x_1'^2+a_{xx}x_1^2+a_{x'x}x_1x_1'+a_{x\delta}x_1\delta+a_{x'\delta}x_1'\delta+a_{y'y'}y_1'^2+a_{yy}y_1^2+a_{\delta\delta}\delta^2, \tag{9}$$

$$y_2''-a_{y'}y_2'-a_yy_2=a_{xy}x_1y_1+a_{y'x}y_1'x_1+a_{y\delta}y_1\delta+2x_1'y_1'+a_{y'\delta}y_1'\delta, \tag{10}$$

$$\tau_2'=c_{x'x'}x_1'^2+c_{xx}x_1^2+c_{x\delta}x_1\delta+c_{x\gamma}x_1\gamma+c_{y'y'}y_1'^2+c_{yy}y_1^2+c_xx_2+c_{\delta\delta}\delta^2+c_{\delta\gamma}\delta\gamma+c_{\gamma\gamma}\gamma^2. \tag{11}$$

Here the coefficients are given by the expressions

$$a_{x'}=a_{y'}=-a_{x'\delta}=-a_{y'\delta}=-\varepsilon^2\frac{sd}{2f},$$

$$a_x=c-2+\varepsilon^2\left(-\frac{5}{6}c^3+\frac{10}{3}c^2-\frac{19}{6}c+\frac{c}{2f^2}-\frac{1}{2f^2}-\frac{10}{3}\right),$$

$$a_\delta=1+\varepsilon^2\left(-\frac{1}{3}c^2+\frac{5}{3}c+\frac{5}{3}\right),$$

$$a_0=\varepsilon^2\left(\frac{1}{3}c^2-\frac{2}{3}c-\frac{2}{3}\right)+\varepsilon^4\left(-\frac{1}{12}c^4+\frac{27}{40}c^3-\frac{29}{36}c^2+\frac{31}{180}\,\frac{c^2}{f^2}-\frac{361}{360}c-\frac{17}{45}\,\frac{c}{f^2}+\frac{37}{180}\,\frac{1}{f^2}-\frac{67}{90}\right),$$

$$a_y=-c+\varepsilon^2\left(-\frac{1}{6}c^3-\frac{1}{6}c^2-\frac{2}{3}c+\frac{c}{2f^2}-\frac{1}{2f^2}\right),$$

$$c_x=2+\varepsilon^2\left(-\frac{1}{3}c^2+\frac{5}{3}c+\frac{5}{3}\right),$$

$$c_{\delta}=-\frac{1}{2}-\frac{3}{4}\varepsilon^{2}(c+1),\quad c_{\gamma}=\frac{1}{2}+\frac{1}{4}\varepsilon^{2}(c+1),$$

$$c_{0}=\frac{1}{2}\varepsilon^{2}(c+1)$$

$$+\varepsilon^{4}\left(-\frac{1}{8}c^{3}+\frac{17}{24}c^{2}+\frac{11}{24}c+\frac{5}{24}\frac{c}{f^{2}}-\frac{5}{24}\frac{1}{f^{2}}+\frac{11}{24}\right),$$

$$a_{xx}=-c^{2}+4c-5+\varepsilon^{2}\left(c^{4}-\frac{71}{12}c^{3}+\frac{37}{3}c^{2}\right.$$

$$\left.-\frac{2}{3}\frac{c^{2}}{f^{2}}-\frac{25}{4}c+\frac{7}{3}\frac{c}{f^{2}}-\frac{5}{3}\frac{1}{f^{2}}-\frac{28}{3}\right),$$

$$a_{x'x}=a_{y'x}=\varepsilon^{2}s\left(-\frac{2}{3}c^{2}+\frac{4}{3}c-\frac{2}{3}\right),$$

$$a_{x'x'}=1+\varepsilon^{2}\left(\frac{1}{3}c^{2}-\frac{2}{3}c-\frac{2}{3}\right),$$

$$a_{x\delta}=-c+5+\varepsilon^{2}\left(\frac{5}{6}c^{3}-\frac{17}{3}c^{2}+\frac{59}{6}c-\frac{c}{2f^{2}}+\frac{1}{2f^{2}}+11\right),$$

$$a_{yy}=c^{2}-\frac{1}{2}c+\varepsilon^{2}\left(\frac{1}{4}c^{4}+\frac{1}{2}c^{3}+\frac{1}{4}\frac{c^{3}d}{f}-\frac{1}{4}c^{2}-\frac{1}{4}\frac{c^{2}d}{f}\right.$$

$$\left.-\frac{7}{12}\frac{c^{2}}{f^{2}}-c+\frac{11}{12}\frac{c}{f^{2}}-\frac{1}{4}\frac{cd}{f^{3}}-\frac{1}{3f^{2}}+\frac{1}{4}\frac{d}{f^{3}}\right),$$

$$a_{y'y'}=-1+\varepsilon^{2}\left(\frac{1}{3}c^{2}-\frac{2}{3}c-\frac{2}{3}\right),$$

$$a_{\delta\delta}=-1+\varepsilon^{2}\left(\frac{1}{3}c^{2}-\frac{8}{3}c-\frac{8}{3}\right),$$

$$a_{xy}=2c^{2}-3c+\varepsilon^{2}\left(\frac{1}{2}c^{4}+\frac{2}{3}c^{3}+\frac{1}{2}\frac{c^{3}d}{f}-\frac{5}{6}c^{2}-\frac{1}{2}\frac{c^{2}d}{f}\right.$$

$$\left.-\frac{7}{6}\frac{c^{2}}{f^{2}}-\frac{10}{3}c+\frac{17}{6}\frac{c}{f^{2}}-\frac{1}{2}\frac{cd}{f^{3}}-\frac{5}{3f^{2}}+\frac{1}{2}\frac{d}{f^{3}}\right),$$

$$a_{y\delta}=c+\varepsilon^{2}\left(\frac{1}{6}c^{3}+\frac{7}{6}c^{2}+\frac{5}{3}c-\frac{c}{2f^{2}}+\frac{1}{2f^{2}}\right),$$

$$c_{xx}=-\frac{1}{2}c+2+\varepsilon^{2}\left(\frac{5}{12}c^{3}-\frac{9}{4}c^{2}+3c-\frac{c}{4f^{2}}+\frac{1}{4f^{2}}+\frac{10}{3}\right),$$

$$c_{x\delta}=-2+\varepsilon^{2}\left(\frac{1}{2}c^{2}-4c-4\right),$$

$$c_{x\gamma}=1+\varepsilon^{2}\left(-\frac{1}{6}c^{2}+\frac{5}{6}c+\frac{5}{6}\right),$$

$$c_{yy}=\frac{1}{2}c+\varepsilon^{2}\left(\frac{1}{12}c^{3}+\frac{1}{3}c^{2}+\frac{7}{12}c-\frac{c}{4f^{2}}+\frac{1}{4f^{2}}\right),$$

$$c_{x'x'}=c_{y'y'}=\frac{1}{2}+\frac{1}{4}\varepsilon^{2}(c+1),\quad c_{\delta\delta}=\frac{3}{8}+\frac{15}{16}\varepsilon^{2}(c+1),$$

$$c_{\delta y}=-\frac{1}{4}-\frac{3}{8}\varepsilon^{2}(c+1),\quad c_{\gamma\gamma}=-\frac{1}{8}-\frac{1}{16}\varepsilon^{2}(c+1),$$

where $c=\cos\varphi/f$ and $s=\sin\varphi/f$.

More cumbersome equations for higher-order aberrations can be obtained analogously. In the limit of a capacitor having an indefinitely narrow interelectrode gap $\varepsilon\to 0$ Eqs. (6) and (7) coincide with the paraxial path equations derived in Ref. 14 on the assumption of a constant potential and a constant field on the axial path of the particle beam.

Linearly independent solutions of the homogeneous linear equations whose left sides coincide with the left sides of Eqs. (6) and (7) can be calculated only by numerical integration. Once this has been done, however, solutions of the inhomogeneous equations (6) and (8)–(11) with zero initial conditions are readily found by the method of variation of arbitrary constants in the form of integrals; these aberration integrals are too cumbersome to write out in explicit form.

We have thus reduced the general solution of the problem of calculating the paths and time of flight to the sum of a combination of solutions of homogeneous linear equations corresponding to the specified initial conditions and of the indicated aberration integrals. Numerical experiments have shown that even when third-order aberrations are taken into account, the time to calculate the properties of the polar toroidal capacitor on the basis of the proposed approach is a fraction of a second on a late-model personal computer, bringing into reality the efficient optimization of devices containing such capacitors, which requires the scanning of a large number of alternative configurations.

We note that Eq. (6) is inhomogeneous even for a main particle with the nominal energy ($\delta=0$) owing to the presence of the coefficient $a_0$. This means that the main particle deviates somewhat from the circular optical axis in the capacitor. For this reason the exact values of the aberration coefficients of a fixed order depend not only on the solutions of the equations corresponding to this order, but also on the higher-order aberration integrals. In particular, the paraxial properties of the capacitor, determined as the solutions of Eqs. (6)–(8), are inaccurately described unless the contributions of the second-order aberration integrals are taken into account.

To estimate these errors and to compare the accuracy of the proposed computational approach with that of numerical simulation methods and previously proposed approximate approaches, we calculate the ion-optical properties of a two-stage analyzer for investigations of the astrophysical plasma;[4] the analyzer geometry is shown in Fig. 2. It consists of two polar toroidal capacitors in series. The radii of the optical axes are $r_1=43$ mm in the first capacitor and $r_2=60$ mm in the second capacitor, corresponding to angles of rotation $\varphi_1=127.6°$ and $\varphi_2=85°$, respectively, and the distances of the centers of meridional curvature of the electrodes from the symmetry axis are $a_1=119$ mm and $a_2=100.2$ mm. The angle of entry of the particle beam into the system relative to the direction of the axis of rotation is

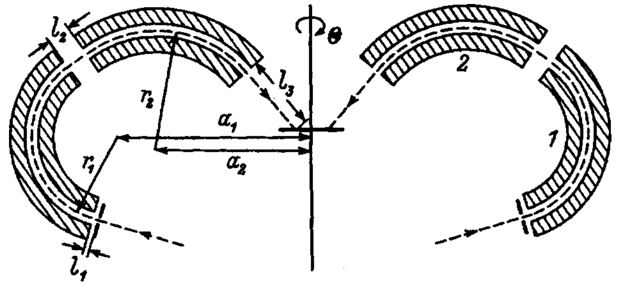

FIG. 2. Relative position of polar toroidal capacitors in a two-stage analyzer. *1*) First stage; *2*) second stage.

$\varphi_0 = 73°$, the distance from the input slit to the first stage is $l_1 = 3$ mm, the spacing of the capacitors is $l_2 = 11$ mm, and the distance from the second stage to the plane of the detector is $l_3 = 37.5$ mm. We calculate the transfer matrix coefficients of the system from the input slit to the detector plane (standard notation such as that used in Ref. 15 are used to represent the transfer matrix coefficients). We note that the investigated configuration is characterized by angular focusing in the radial direction [i.e., the transfer matrix coefficient $(x|\alpha) = 0$, where $\alpha$ is the angle of the path relative to the optical axis in the radial direction) and by zero spatial dispersion with respect to energy, $(x|\delta) = 0$.

We have checked the accuracy of the proposed algorithm by comparing the first-, second-, and third-order transfer matrix coefficients of the analyzer as determined from calculations of the aberration integrals, and also by highly accurate numerical integration of the paths using methods of differential algebra.[16] Our test shows that when the contributions of aberration integrals up to and including third-order are taken into account, the first- and second-order aberration coefficients are calculated within 0.1% error limits, and the third-order aberration coefficients are calculated within 10% error limits. The large error of the third-order aberrations is attributable to the rejection of higher-order aberration integrals, which, as explained above, are required in order to attain higher accuracy.

Table I shows a comparison of the first-order transfer matrix coefficients, calculated by the method proposed in the article with the contributions of aberration integrals up to and including third-order taken into account (column AI), by the approximate method proposed in Ref. 4 (column AM), and by numerical calculations of the paths (column NC). The results in the last two columns are taken from Ref. 4; the values of the coefficients are given in the dimensionless quantities used in the present study, where the initial coordinate is expressed in units of the radius of the axis of the first stage, and the final coordinate is expressed in units of the radius of the second stage [so that, in particular, the product of linear $(x|x)$ and angular $(\alpha|\alpha)$ magnifications in the table are not equal to unity]; angular variables are expressed in the tangents of angles, energy and mass deviations are expressed in units of their nominal values, and the time variable is expressed in the product of the relative deviation of the time of flight from its nominal value and the total length of the optical axis divided by the radius of the second stage. It is evident from Table I that our results agree significantly better with the results of the numerical calculations than do the results obtained by the approximate method in Ref. 4 with the one exception of the coefficient $(x|x)$, where the discrepancies are of the same order of magnitude. We note that the discrepancy between our results and the numerical simulation results are of the order of one percent, which is not as good as the above estimate of the error of the method of aberration integrals. The reason is that our calculations have ignored edge effects at the boundaries of the capacitors.

TABLE I. Transfer matrix coefficients of a two-stage analyzer.

| | AI | AM | NC |
|---|---|---|---|
| $(x\|x)$ | 0.875 | 0.867 | 0.870 |
| $(\alpha\|x)$ | 1.124 | 1.047 | 1.110 |
| $(\alpha\|\alpha)$ | 0.819 | 0.827 | 0.821 |
| $(\alpha\|\delta)$ | −0.690 | −0.656 | −0.680 |
| $(\tau\|x)$ | −1.246 | −1.138 | −1.234 |
| $(\tau\|\delta)$ | 2.000 | 1.880 | 1.991 |
| $(\tau\|\gamma)$ | 1.966 | 1.954 | 1.967 |

TABLE II. Comparison of calculated and measured transfer matrix coefficients for various positions of the plane of observation.

| | Calculated | | | Measured | | |
|---|---|---|---|---|---|---|
| | I | II | III | I | II | III |
| $(x\|x)$ | −1.31 | 0.89 | 1.38 | −1.30 | 0.90 | 1.40 |
| $(x\|\alpha)$ | −0.78 | 0.00 | 0.34 | −0.86 | −0.08 | 0.38 |
| $(x\|\delta)$ | 1.35 | −0.04 | −0.38 | 1.33 | −0.17 | −0.37 |

Table II compares experimentally measured first-order transfer matrix coefficients with those calculated by the method of aberration integrals for a two-stage system whose geometry is given in Ref. 7. The relative position of the stages in this system is the same as in Fig. 2, but the geometrical parameters differ somewhat from those described in Ref. 4. For the given system $l_1 = 6$ mm, $r_1 = 44$ mm, $a_1 = 104$ mm, $\varphi_0 = 76°$, $\varphi_1 = 122.5°$, $l_2 = 19$ mm, $r_2 = 51$ mm, $a_2 = 85.4$ mm, and $\varphi_2 = 86.3°$. The measurement results are taken from Table 2 in Ref. 7. In our Table II column I gives the coefficients measured at a distance of 40 mm after the first capacitor (without the second one), column II gives the coefficients measured at a distance of 24 mm after the second capacitor, and column III gives the coefficients at a distance of 42 mm after the second capacitor. A comparison of the computational and measurement results indicates good agreement between them.

The proposed method therefore has the high accuracy and economy required for optimization calculations of the paraxial and aberration properties of polar toroidal charged-particle beam analyzers.

This work has received partial support from International Science Foundation Grant No. R6F300.

[1] R. G. Smeenk, R. M. Tromp, H. H. Kersten *et al.*, Nucl. Instrum. Methods **195**, 581 (1982).

[2] H. Engelhardt, W. Baeck, D. Menzel, and H. Liebl, Rev. Sci. Instrum. **52**, No. 835 (1981).

[3] R. C. J. Leckey, J. Electron Spectr. **43**, 183 (1987).

[4] A. G. Ghielmetti and E. G. Shelley, Nucl. Instrum. Methods Phys. Res. A **298**, 181 (1990).

[5] D. T. Young and J. A. Marshall, Nucl. Instrum. Methods Phys. Res. A **298**, 227 (1990).

[6] E. Moebius, P. Bochsler, A. G. Ghielmetti, and D. C. Hamilton, Rev. Sci. Instrum. **61**, No. 3609 (1990).

[7] O. Bratschi, A. G. Ghielmetti, E. G. Shelley, and H. Balsiger, Rev. Sci. Instrum. **64**, No. 184 (1993).

[8] H. Moestue, Rev. Sci. Instrum. **44**, No. 1709 (1973).

[9] V. T. Kogan and A. K. Pavlov, Zh. Tekh. Fiz. **55**, 2079 (1985) [Sov. Phys. Tech. Phys. **30**, 1225 (1985)].

[10] M. I. Yavor, H. Wollnik, M. Nappi, and B. Hartmann, Nucl. Instrum. Methods Phys. Res. A **311**, 448 (1992).

[11] R. Tofoletto, R. C. J. Leckey, and J. D. Riley, Nucl. Instrum. Methods Phys. Res. A **12**, 282 (1985).

[12] O. A. Abaza and M. I. Yavor, Nauchn. Priborostr. **1**, 179 (1991).

[13] H. Matsuda, Nucl. Instrum. Methods **91**, 637 (1971).

[14] V. V. Zashkvara, M. I. Korsunkiĭ, and V. S. Red'kin, Zh. Tekh. Fiz. **38**, 1336 (1968) [Sov. Phys. Tech. Phys. **13**, 1092 (1968)].

[15] H. Wollnik, *Optics of Charged Particles*, Academic Press, Orlando (1987) [Russian trans., Énergoatomizdat, Leningrad, 1992].

[16] M. Berz, IEEE Trans. Electron Devices **ED-35**, 2002 (1988).

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Possibility of lowering the effective emittance of neutralized ion beams

S. Yu. Udovichenko

*D. V. Efremov Scientific-Research Institute of Electrophysical Apparatus, 189631 St. Petersburg, Russia*

(Submitted November 11, 1996)
Zh. Tekh. Fiz. **68**, 112–115 (April 1998)

Self-similarity conditions are determined for the steady states of a quasineutral beam plasma generated during the transport of ion beams in a gaseous medium. The self-consistent radial distribution of the beam and plasma densities is taken into account. Under the resulting conditions the radial electric field of the beam–plasma system is linear, and it is possible for beams to be transported without nonlinear distortion of their phase response characteristics or an increase in the effective emittance. [S1063-7842(98)02004-2]

## INTRODUCTION

The transport of an ion beam in a gas flowing into an injector from a plasma source is accompanied by the buildup of secondary charged particles formed as neutral atoms are ionized by the beam. The space charge of fast ions is partially or fully neutralized, depending on the parameters of the beam–plasma system that is formed. The resulting beam–plasma field can influence the dynamics of precision beams in injection devices.

The static electric field has been determined for fully charge-neutralized beams in rarefied and dense gaseous media[1–3] and also for deneutralized beams.[4]

Numerical modeling has been used[3,5,6] to investigate the influence of the evaluated static electric field of a quasineutral beam–plasma system on the dynamics of a precision ion beam with a small phase volume and small angular divergence. It has been shown that when the beam is transported through a ''plasma lens,'' its phase volume acquires nonlinear distortions, and its effective emittance increases considerably. A convergent beam can be obtained from a divergent beam, and vice versa, depending on the choice of parameters of the plasma lens. The results are used to optimize the parameters of precision beams in the interest of matching injectors with an high-frequency accelerator and in connection with long-range beam transport in a gaseous medium.

The cause of the distortion of the phase volume of the beam and the attendant increase in its effective emittance lies in the nonlinearity of the forces produced by the resultant electric field of the quasineutral or charged beam plasma. In the presence of a plasma medium a self-consistent motion of the beam particles takes place in the field of the underneutralized space charge produced by the fast particles themselves and by secondary charged particles. In previous determinations[1–3] of the electric field of a beam plasma the motion of the beam particles has been assumed not to be self-consistent, because the authors used the approximation of a specified beam density profile. This approximation can become non-self-similar, and the motion of the beam takes place under the influence of nonlinear forces produced by the electric field.

Goncharov[7] has demonstrated the possibility of self-similar spreading of a neutralized beam of positive ions under the influence of linear forces produced by the electric field of the beam plasma. In this case the radial spreading of the beam is such that each new state of the beam is a function of time only and not of the preceding state, i.e., it is self-similar. However, the approximation used in Ref. 7 — a self-consistent beam density profile and a specified plasma density profile — is crude and yields two results that conflict with experimental data. Under conditions such that the densities $n_i$ and $n_e$ of the ionic and electronic components of the plasma greatly exceed the beam density $n_b^+$, a solution is obtained for the potential drop across the beam radius $\Delta\varphi=(n_b^+/n_e)T_e/e$, where $T_e$ is the plasma electron temperature. But the experimental value is close to the positive ambipolar potential drop of a quasineutral plasma ($n_e\approx n_i \gg n_b^+$): $\Delta\varphi\leqslant T_e/e$. It follows from the solution of the equation for a self-consistent potential in Ref. 7 that the potential drop in the volume of a beam of negative ions should be negative. It is a well-known fact that in a dense gas this potential difference is positive and of the same order of magnitude as the potential difference in the case of a beam of positive ions.[8] The incorrectness of the approximation is attributable to the underlying assumption that the slow secondary electrons are not perturbed by the electric field of the beam plasma. It will be shown below that the density profile of the plasma ions significantly influences the choice of self-similar solution. Moreover, in the cited investigation the density of secondary electrons is a given parameter, and the plasma generation mechanism is not determined.

The objective of the present study is to determine the conditions for linearity of the electric field (self-similarity of states) of a quasineutral beam plasma, when it is then possible to lower the effective emittance of ion beams. Here we use the approximation of self-consistent density profiles for the beam and the plasma. We consider a broad range of variation of the gas pressure, when the beam-generated plasma can be collisional or collisionless. Together with beams of positive ions we investigate the self-similar states of beams of negative ions.

## SELF-SIMILAR STATE OF A DENSE BEAM PLASMA

We consider the plasma generated by a beam during the ionization of a dense gaseous medium, in which the mean free path of a slow ion in relation to elastic collisions with neutral atoms is shorter than the radial dimension of the ion conductor ($\lambda_{i0}<R$). To determine the spatial structure of the collisional plasma and its ambipolar field, it is necessary to use the system of hydrodynamic equations of motion and continuity of slow charged particles, along with the energy balance equation for the plasma electrons. This closed system of equations has been investigated previously[1] for the case of a dense plasma ($n_e$, $n_i \gg n_b^{\pm}$). The electric field does not appear explicitly in the ambipolar diffusion equation obtained from the equations of continuity and motion of the plasma components. The problem is therefore separable. The boundary condition on the potential is not needed for determining the density profile of the plasma.

By analogy with Ref. 7 we assume that a self-similar state of a beam of positive and negative ions is satisfied by a parabolic radial distribution of the density of fast particles: $n_b^{\pm}(\xi)=n_b^{\pm}(0)(1-C\xi^2)$, where $\xi=r/r_b$, $r_b$ is the radius of the cylindrical beam, and $C$ is a constant. We use this beam density distribution to describe the plasma particle-generating source in the ambipolar diffusion equation:[1]

$$\frac{1}{\xi}\frac{d}{d\xi}\left[\xi D_A \frac{dn_i}{d\xi}\right]=-\nu_H n_b^{\pm}(\xi) r_b^2, \tag{1}$$

where $D_A=2T_e/m_i\nu_{i0}$, $\nu_{i0}$ is the frequency of elastic collisions of slow ions with gas atoms, $\nu_H=\sigma_i n_g v_b$ is the frequency of formation of plasma particles, $\sigma_i$ is the cross section of ionization of a gas atom by a beam ion, and $v_b$ is the velocity of the beam.

The solution of Eq., (1) has the form

$$n_i(\xi\leqslant 1)=n_i(1)+an_b^{\pm}(0)\left[1-\frac{C}{4}-\xi^2\left(1-\frac{C\xi^2}{4}\right)\right],$$

$$n_i(\xi\geqslant 1)=n_w-2an_b^{\pm}(0)\ln(r/R), \tag{2}$$

where $n_i(1)=n_w+2an_b^{\pm}(0)\ln(R/r_b)$, $n_w=(1-C/2)\times\nu_H r_b^2 n_b^{\pm}(0)/2Rv_s$ is the boundary density of a quasineutral plasma, $a=\nu_H r_b^2/4D_A$, $R$ is the radius of the sheath at which the plasma recombines, and $v_s=(T_e/m_i)^{1/2}$ is the velocity of the ion beam.

In determining the effective boundary condition for a quasineutral plasma at the boundary of the charged layer for a nonconducting sheath, we have assumed that a diffusion regime prevails in the volume of the plasma, and the charged layer is collisionless. An appropriate boundary condition must be used for a collisional charged layer.[1]

We write the equation for a self-consistent potential of the electric field of the beam plasma for a Boltzmann distribution of secondary electrons in the form

$$\frac{1}{\xi}\frac{d}{d\xi}\,\xi\frac{d\varphi}{d\xi}=-\frac{A}{n_b^{\pm}(0)}$$

$$\times[n_i(\xi)\pm n_b^{\pm}(\xi)-n_e(1)\exp(e\varphi/T)_e], \tag{3}$$

where $A=4\pi e n_b^{\pm}(0) r_b^2$.

The condition of equal densities of the components of the unperturbed plasma is attained for $\xi\geqslant 1$: $n_i(1)\approx n_e(1)$. By analogy with Ref. 7 in this case we have $\varphi=\alpha(1-P\xi^2)/4$, where $\alpha$ and $P$ are constants characterizing the unneutralized charge. Setting $e\varphi \ll T_e$ in Eq., (3), we obtain the following expressions for the unknown constants:

$$\alpha=4aAd_e^2/Pr_b^2, \quad P=(1\pm C/a)(1-C/4)^{-1}, \tag{4}$$

where the plus and minus signs correspond to beams of positive and negative ions, respectively, $d_e=[T_e/4\pi e^2 n_e(1)]^{1/2}$ is the Debye screening length of the electrons, $a\gg 1$, and $4d_e^2/r_b^2 \ll 1$.

According to Eq., (4), solutions decaying away from the beam axis for a self-consistent potential and density of fast particles is obtained for $0\leqslant C\leqslant 1$, $3/4\leqslant P\leqslant 1$. For example, in the case of a beam with a uniform particle density ($C=0$, $P=1$) the self-similar solution for the potential has the form

$$\varphi(\xi)=\frac{T_e}{e}\frac{n_b^{\pm}(0)}{n_e(1)}a(1-\xi^2), \tag{5}$$

where, for example, if $n_e(1)\approx n_i(1)\gg n_w$, we can set $an_b^{\pm}(0)/n_e(1)\approx 1/2\ln(R/r_b)$.

It is evident from Eq., (5) that the approximation $\varphi \ll T_e/e$ is well satisfied near the surface of the beam, where the field is a maximum and where it was previously observed to be highly nonlinear. Near the beam axis the field tends to zero, and the error of determination of the quantity $\varphi(0)$ is insignificant.

If, following Ref. 7, we set $n_i(0)\approx n_e(0)$ and $n_e(\xi)=n_e(0)\exp(e\varphi/T_e)$ in (3), where $\varphi$ is determined to within a constant, for the density of a beam of positive ions we obtain a growing solution $C\approx 1-a$, where $a\gg 1$. The fact is that such a crude quasineutrality condition cannot be admitted, because $n_i(0)-n_e(0)\approx n_b^{+}(0)$.

It is necessary that the potential distribution in the beam (5) be matched with the potential distribution of the surrounding plasma. The potential in the region of the beam is determined to within an arbitrary constant $C_1$. It follows from the matching of the solutions at the boundary of the beam $\xi=1$ that to disregard $C_1$ for $r_b\gg d_e$ is justified.[7]

## CONDITIONS FOR LINEARITY OF THE FIELD IN A RAREFIED BEAM PLASMA

Here we consider the plasma generated by an ion beam during the ionization of a rarefied gas, when the reverse condition $\lambda_{i0}>R$ holds. For a low gas pressure, if $A_0=\nu_H r_b/2v_s \ll 1$, the following quasineutral state is established in a beam of positive ions: $n_b^{+}+n_i\approx n_e$; $n_i \ll n_b^{+}$, $n_e$. According to Eq., (3), this steady state of the beam plasma is self-similar for a Gaussian distribution of the beam density:

$$n_b^{+}(\xi)\approx n_e(\xi)=n_b^{+}(0)\exp(-\xi^2),$$

$$\varphi(\xi)\approx(T_e/e)(1-\xi^2). \tag{6}$$

However, this approximation fails at a certain distance from the beam axis, because the density of fast ions is comparable with the density of slow ions $n_i$. For $\xi \gg 1$ the opposite condition holds: $n_i \approx n_e \gg n_b^+$.

We now determine the region of the beam in which the approximation (6) is valid and in which the electric field is a linear function of the coordinate. We find the radial distribution of the density of slow ions by means of the equation of motion[2,3] and the expression for the plasma flux, which follows from the equation of continuity:

$$u_i \frac{du_i}{d\xi} = -v_S^2 \frac{d}{d\xi} \ln(n_e) - \frac{S u_i r_b}{n_i}, \tag{7}$$

$$n_i u_i = A_0 n_b^+(0) v_S \xi^{-1} (1 - f(\xi)), \tag{8}$$

where $S = \nu_H n_b^+$ is the source of enhanced ionization of the plasma, and $f(\xi) = \exp(-\xi^2)$.

The solution has the form

$$\frac{n_i(\xi)}{n_b^+(0)} = \frac{A_0}{2^{1/2}\xi} (1 - f(\xi))^2 [\xi^2 - 2(1 - f(\xi)) + 2^{-1}(1 - f^2(\xi))]^{-1/2}. \tag{9}$$

An analysis of Eq., (9) shows that for $A_0 = 10^{-1}$ the ratio $n_i / n_b^+$ has the value $2 \times 10^{-1}$ at the point $\xi = 1$ and the value $5 \times 10^{-1}$ at the point $\xi = 1.5$. The coordinate $\xi = 1$ is the dimensionless effective radius of a beam with a Gaussian particle density profile, which takes in 63% of the particles. The percentage of particles taken in at $\xi = 1.5$ is 90%. Consequently, when the gas pressure is lowered ($A_0 = 2 \times 10^{-2}$), 90% of the beam particles can be transported under conditions of a plasma electric field that is linear in the transverse coordinate ($n_i n_b^+ = 10^{-1}$ at $\xi = 1.5$).

The following quasineutrality condition holds in the case of a beam of negative ions at a low gas pressure ($A_0 \ll 1$): $n_b^- + n_e \approx n_i$; $n_e \ll n_b^-$, $n_i$. There are few electrons in the volume of a neutralized ion beam, but the electric field depends specifically on this highest-mobility component of the plasma. It follows from Eq., (7) with $S$ set equal to $\nu_H n_b^-$ that for a Boltzmann distribution of electrons and a linear electric field the directional velocity of slow ions is also a linear function of the transverse coordinate. According to the equation of continuity, for the ionic plasma component this condition is met for a uniform beam density profile and $u_i(\xi) \approx A_0 v_S \xi$. The corresponding self-similar solution for the potential of a stable neutralized beam of negative ions,

$$\varphi(\xi) = \frac{T_e}{e} \ln(n_e) \approx \frac{3}{2} \frac{T_e}{e} A_0^2 (1 - \xi^2), \tag{10}$$

agrees with the field determined in a numerical calculation.[2]

## CONCLUSION

Self-similar states of a quasineutral beam plasma generated during the transport of stable ion beam in a gaseous medium have been found by solving the equations of continuity and motion for secondary charged particles and the Poisson equation for a self-consistent potential of the electric field. The proposed approach eliminates the shortcomings of existing models of the steady state of a beam plasma, where a less than fully self-consistent problem is solved and the potential of the electric field is sought for a specified distribution of the density of the beam or the plasma.

Conditions have been determined for self-similarity of the steady states of neutralized beams of positive and negative ions, when the plasma produced by them can be collisional or collisionless with respect to the ionic component. Under these conditions the radial electric field of the beam–plasma system is a linear function of the transverse coordinate, i.e., it is possible for beams to be transported without nonlinear distortion of the phase response characteristics and an increase in the effective emittance.

Collective processes associated with the excitation of the spectra of eigenmodes of the plasma have been disregarded in the search for self-similar states of neutralized beams. In an unstable beam fully developed plasma oscillations lead to collective heating of the beam particles, an increase in the effective emittance, and dynamic decompensation of the space charge of the beam. The possibility of diminishing collective heating and thereby lowering the effective emittance of the beam has been investigated previously.[9] The conditions under which steady states of unstable, partially neutralized beams of positive and negative ions are possible have also been determined.[4]

The electron temperature in Eqs. (5), (6), and (10) is determined on the basis of the solution of the energy balance equation derived in Ref. 1. Together with collisionless (collective) and collisional heating of electrons in the field of plasma oscillations, other heating sources and cooling mechanisms over a wide range of gas and beam densities are also taken into account in this equation.

Experimental data[8] have shown that for a low gas pressure in an unstable neutralized beam of negative ions in the presence of fully developed plasma oscillations the potential drop across the radius of the beam is much higher than the potential drop determined from Eq., (10) and is of the same magnitude as the potential drop in a neutralized beam of positive ions.

[1] V. Yu. Udovichenko, Fiz. Plazmy **22**, 149 (1996) [Plasma Phys. Rep. **22**, 137 (1996)].

[2] E. B. Hooper, O. A. Andersen, and P. A. Willmann, Phys. Fluids **22**, 2334 (1979).

[3] P. N. Afanas'ev, Yu. A. Svistunov, V. P. Sidorov, and S. Yu. Udovichenko, NIIÉFA Preprint No. P-V-0759 [in Russian], (D. V. Efremov Scientific-Research Institute of Electrophysical Apparatus, Moscow, 1987).

[4] S. Yu. Udovichenko, Zh. Tekh. Fiz. **64**(8), 104 (1994) [Tech. Phys. **39**, 802 (1994)].

[5] V. P. Sidorov, S. Yu. Udovichenko, A. M. Astapkovich *et al.*, in *Proceedings of the Symposium on the Production and Neutralization of Negative Ions and Beams* (Brookhaven, USA, 1989), pp. 614–628.

[6] P. N. Afanas'ev, Yu. A. Svistunov, V. P. Sidorov, and S. Yu. Udovichenko, Vopr. At. Nauki Tekh. Ser. Termoyad. Sintez, No. 1, 28 (1989).

[7] A. A. Goncharov, Ukr. Fiz. Zh. **33**, 371 (1988).

[8] J. Sherman, E. Pitcher, R. Stevens, and P. Allison, in *Proceedings of the Symposium on the Production and Neutralization of Negative Ions and Beams* (Brookhaven, USA, 1992), pp. 686–694.

[9] V. P. Sidorov and S. Yu. Udovichenko, Vopr. At. Nauki Tekh. Ser. Termoyad. Sintez, No. 4, 31 (1992).

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Formation of relativistic positron systems by the axial channeling of positrons in ionic crystals

A. S. Gevorkyan and A. G. Grigoryan

*Institute of High-Speed Computations and Data Bases of the State Committee of the Russian Federation on Science and Technologies, 194291 St. Petersburg, Russia*

A. R. Mkrtchyan and A. G. Toneyan

*Institute of Applied Problems in Physics, Academy of Sciences of Armenia, 375014 Erevan, Armenia*

(Submitted December 26, 1996)
Zh. Tekh. Fiz. **68**, 116–120 (April 1998)

An analytical expression is written for the effective interaction potential of a fast charged particle with ionic crystals of the CsCl type as a function of the temperature of the medium, taking long-range order into account. A numerical analysis shows that there is a real possibility of axial superchanneling of positrons along the ⟨100⟩ axis of negatively charged ions in crystals with this structure. The wave function and energy spectrum of the localized state are investigated, and the possibility of the formation of metastable, two-dimensional, relativistic positron systems is analyzed. © *1998 American Institute of Physics.* [S1063-7842(98)02104-7]

## INTRODUCTION

The phenomenon of anomalous transmission of ions along definite crystallographic axes and planes was discovered experimentally in 1960.[1] In 1963 it was confirmed by numerical simulation[2] and became known as the channeling effect. In 1965 Lindhard published a theoretical explanation of the phenomenon in the framework of classical mechanics.[3] The quantum mechanical theory of electron and positron channeling has been elaborated by many authors.[4–6] The years following the publication of Lindhard's theory saw the burgeoning of theoretical and experimental work on the problems of channeling of light particles — electrons and positrons.

It is important to note that an electron in a crystal can undergo both planar and axial channeling. So far only one type of pure channeling is known for positrons: the regime where a particle is localized between two adjacent planes.[5]

The feasibility of axial channeling of positive particles has not been give serious consideration to date, because the crystallographic axes themselves are positively charged, regardless of the species. At the same time, to explore the possibilities of axial channeling of positrons and, hence, the formation of metastable relativistic positron systems pose a problem of utmost urgency in radiation physics. Suffice it to say that currently one of the possible approaches to the generation of coherent $\gamma$ rays is seen in the method of stimulation of decay of positrons $(e^+e^-)\rightarrow 2\gamma$ formed by the displacement of relativistic electron and position beams in vacuum.[7,8] However, this approach is extremely tenuous at best, specifically in light of the low probability of generation of the $(e^+e^-)$ pairs themselves under the stated conditions. A different method for the generation of coherent $\gamma$ rays has been proposed recently, based on the stimulation of annihilation of positrons channeled in a crystal with host electrons.[9] We believe this to be a preferable approach.

In earlier studies[10–13] we have focused attention on the capabilities of ionic crystals of the CsCl type with a view toward uncovering new possibilities for the channeling of light charged particles. In particular, we have studied in detail the effective interaction potential of a charged particle with a crystal under conditions of planar channeling along the principal ⟨100⟩ planes of cesium $\mathrm{Cs}^+$ and chlorine $\mathrm{Cl}^-$ ions.

In the present article we construct the effective interaction potential of a charged relativistic particle with a crystal near the axial direction ⟨100⟩. We investigate in detail the effective positron channeling potential near the ⟨100⟩ axis of $\mathrm{Cl}^-$ ions by numerical simulation. We solve the Schrödinger equation in a two-dimensional effective potential; we also determine the wave function and the energy spectrum. We analyze the role of dissipative processes in the broadening of the spectral lines of two-dimensional relativistic positron systems.

## POSITRON CHANNELING NEAR THE ⟨100⟩ AXIS OF $\mathrm{Cl}^-$ IONS IN A CsCl CRYSTAL

In Cartesian coordinates with the origin at a designated $\mathrm{Cl}^-$ crystal the potential generated by a three-dimensional, unbounded crystals of the CsCl type in the Jensen–Mayer–Gosler–Rode approximation with allowance for thermal vibrations has the form[10]

$$\varphi(\mathbf{r};T)=4\pi\int d\mathbf{R}\frac{1}{d^3}\sum_{k\neq 0}\frac{1}{k^2}e^{i\mathbf{k}(\mathbf{r}-\mathbf{R})}\times\left\{(-1)^{l+n+m}e^{-\frac{k^2u_{0+}^2}{2}}W^+(\mathbf{R})+e^{-\frac{k^2u_{0-}^2}{2}}W^-(\mathbf{R})\right\},$$

$$\mathbf{k}=\frac{2\pi}{d}(\hat{x}l+\hat{y}n+\hat{z}m), \tag{1}$$

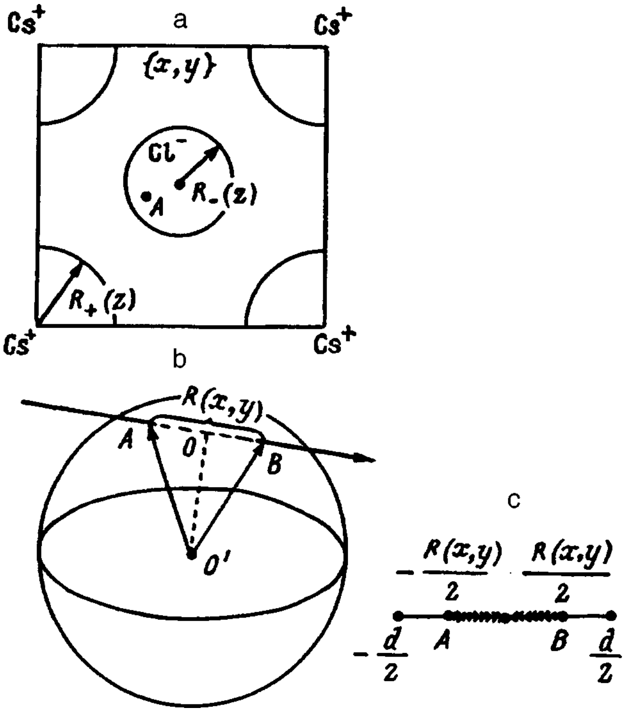

FIG. 1. a) Two-dimensional cross section $\{x,y\}$ of the three-dimensional unit cell of a CsCl crystal at depth $z$ along the $\langle 100\rangle$ axis. The radii of the cross section of the ion spheres are given by the expressions $R_+(z)=\mathrm{Re}\sqrt{R_{0+}^2-(\frac{1}{2}d-z)^2}$ and $R_-(z)=\mathrm{Re}\sqrt{R_0^2-z^2}$; b) Intersection of the particle path with the sphere of a lattice ion at points A and B; c) the hatched part indicates the path segment of length $R(x,y)$ traversed by the particle inside the ion.

where $d$ is the period of the real lattice, $\mathbf{k}$ is the reciprocal lattice vector, $(l,n,m)\in(-\infty,\infty)$ are integers, $u_{0+}\equiv u_{0+}(T)$ and $u_{0-}\equiv u_{0-}(T)$ characterize the amplitudes of thermal vibrations of positive $\mathrm{Cs}^+$ and negative $\mathrm{Cl}^-$ ions at the temperature $T$, and $W^+(\mathbf{R})$ and $W^-(\mathbf{R})$ denote the density of charges in ions of the corresponding signs at the temperature $T=0$.

The electrostatic field (1) is greatly simplified outside the ionic lattice, assuming the form

$$\varphi_{nst}(\mathbf{r};T)=\frac{4\pi e^-}{d^3}\times\sum_{k\neq 0}\frac{1}{k^2}e^{i\mathbf{k}\cdot\mathbf{r}}\left[(-1)^{l+n+m}e^{\frac{-k^2u_{0+}^2}{2}}-e^{\frac{-k^2u_{0-}^2}{2}}\right]. \tag{2}$$

We now investigate the structure of the effective potential. Let a fast, positively charged particle be scattered at a small angle $\vartheta\leqslant\vartheta_L\sim\sqrt{D_0/E}$ (where $\vartheta_L$ is the Lindhard angle, $E$ is the total energy of the particle, and $D_0$ is the depth of the well) on the $\langle 100\rangle$ axis of $\mathrm{Cl}^-$ ion. The potential (1) can then be averaged along the direction of fast motion, i.e., along the $\langle 100\rangle$ axis of $\mathrm{Cl}^-$ ions, which is equivalent to integration of the potential along the coordinate $z$ within the limits of one period $d$ (Fig. 1). If the particle intercepts a Cl $^-$ ion at a distance $\rho=\sqrt{x^2+y^2}$, the segment $d$ is divided into three parts (Fig. 1). In the first and third parts the particle moves outside the ion, and in the second part it moves inside the ion. Note that the length of the path traversed by the particle inside the ion is

$$R(x,y)=2\mathrm{Re}\sqrt{R_0^2-[\eta^2(x)+\eta^2(y)]d^2},$$
$$\eta(x)=\frac{1}{2}+(-1)^P\left(\left\{\frac{x}{d}\right\}-\frac{1}{2}\right),\quad P_x=\left[\frac{2x}{d}\right],$$
$$\eta(y)=\frac{1}{2}+(-1)^{P_y}\left(\left\{\frac{y}{d}\right\}-\frac{1}{2}\right),\quad P_y=\left[\frac{2y}{d}\right], \tag{3}$$

where the brackets $[\ldots]$ and braces $\{\ldots\}$ denote the the integer part and fractional part of a function, respectively. When the particle intercepts a $\mathrm{Cs}^+$ ion, the path is given by the formula

$$R_+(x,y)=2\mathrm{Re}\sqrt{R_{0+}^2-\left[\eta^2\left(x-\frac{1}{2}\right)+\eta^2\left(y-\frac{1}{2}\right)\right]d^2}. \tag{4}$$

We recall that in Eqs. (3) and (4) the symbols $R_{0+}$ and $R_{0-}$ denote the radii of the corresponding ions. We can now write the potential averaged along the coordinate $z$:

$$\varphi_{\mathrm{eff}}(x,y;T)=\int_{-d/2}^{-R_-/2}\varphi_{nst}(\mathbf{r};T)dz+\int_{-R_-/2}^{R_-/2}\varphi(\mathbf{r};T)dz$$
$$+\int_{-R_-/2}^{d/2}\varphi_{nst}(\mathbf{r};T)dz-\int_{-d/2}^{-R_+/2}\varphi_{nst}(\mathbf{r};T)dz$$
$$-\int_{-R_+/2}^{R_+/2}\varphi(\mathbf{r};T)dz-\int_{-R_+/2}^{d/2}\varphi_{nst}(\mathbf{r};T)dz. \tag{5}$$

The substitution of Eqs. (1) and (2) into (5) and elementary integration yield

$$\varphi_{\mathrm{eff}}(x,y;T)=\frac{8e^-}{\pi^2 d}\sum_{\substack{l,n,m=0\\ l+n+m>0}}a_la_na_m\frac{e^{-\lambda^2\mu^2}}{m\mu^2}\cos\left(\frac{2\pi}{d}lx\right)$$
$$\times\cos\left(\frac{2\pi}{d}ny\right)\{(-1)^{l+n}\sin[\pi m\overline{R}_+(x,y)]$$
$$\times W^+(l,n,m)+\sin[\pi m\overline{R}_-(x,y)]$$
$$\times W^-(l,n,m)\}+\frac{4|e^-|}{\pi d}\sum_{\substack{n,l=0\\ n+l>0}}a_la_n\frac{e^{-\lambda^2\nu^2}}{\nu^2}$$
$$\times\cos\left(\frac{2\pi}{d}lx\right)\cos\left(\frac{2\pi}{d}ny\right)\{(-1)^{l+n}-1\},$$
$$\overline{R}_+(x,y)=R_+(x,y)/d,\quad \overline{R}_-(x,y)=R_-(x,y)/d, \tag{6}$$

where

$$\mu^2=l^2+n^2+m^2,\quad \nu^2=l^2+n^2,$$
$$\lambda\equiv u_{0+}/d=u_{0-}/d,\quad a_0=1/2,\quad a_i=1(i\neq 0),$$
$$W^{\pm}(I,n,m)=\int W^{\pm}(\mathbf{R})e^{-i\mathbf{k}\cdot\mathbf{R}}d\mathbf{R}. \tag{6'}$$

We note that the parameter $\lambda$ is obtained on the assumption that the thermal vibration amplitudes are equal to $u_{0+}$

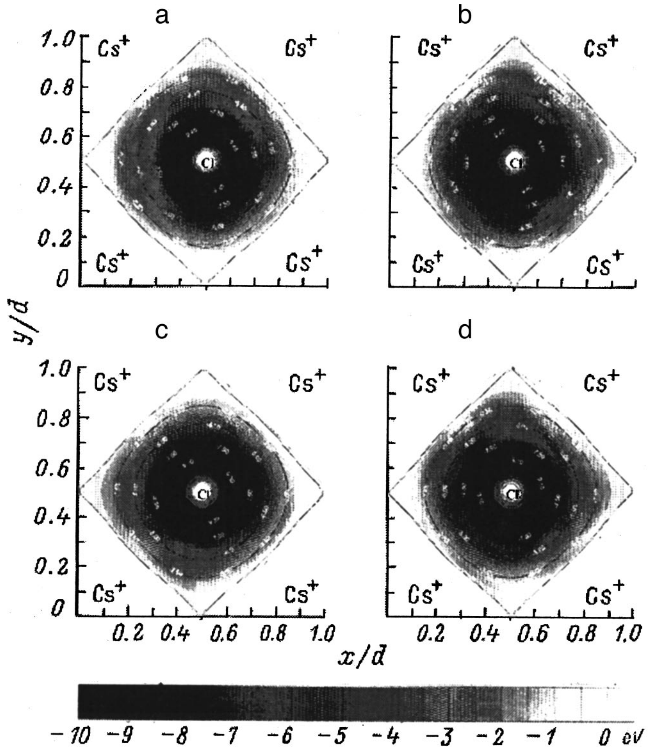

FIG. 2. Profile of the effective potential of axial channeling of a positron along the ⟨100⟩ axis of Cl⁻ ions at various temperatures. a) λ = 0.001; b) 0.01; c) 0.05; d) 0.1.

$=u_{0-}$, which is well within reason for acoustic vibrations. The form of the structure functions $W^{+}(l,n,m)$ and $W^{-}(l,n,m)$ need to be refined for the numerical analysis of Eq. (6). We write the charge density inside the ion in the form

$$W^{\pm}(\mathbf{R})=V^{\pm}(\mathbf{R})+Z^{\pm}\delta(\mathbf{R}),\tag{7}$$

where $V^{\pm}(\mathbf{R})$ is the distribution of electrons inside ions of the crystal, and $Z^{\pm}$ is the number of protons in the point nucleus.

Substituting Eq. (7) into (6′) and assuming that the distribution of the electron charge inside the ion has spherical symmetry, we obtain the following expression for the structure factor:

$$W^{\pm}(l,n,m)=Z^{\pm}+X^{\pm}(l,n,m)\pm 1,$$

$$X^{\pm}(l,n,m)=\frac{4\pi}{k}\int_0^{\infty}V^{\pm}(R)R\sin(kR)\,dR,\quad k=|\mathbf{k}|.\tag{8}$$

Next, adopting the Lenz–Jensen model[10,11] with the parameters of the CsCl crystal in the role of the function $V(R)$, we calculate Eq. (6) for four different values of $\lambda=\{0.001,0.01,0.05,0.1\}$. It is evident from Figs. 2a–2d that a rather broad (of width $\Delta d\sim 0.25d$) potential well exists for fast, positively charged particles around the ⟨100⟩ axis of Cl⁻ ions; the depth of the well is $D_0=9.8$ eV, which remains constant over a broad range of thermal vibrations $0.001\leqslant\lambda\leqslant 0.1$, i.e., over a broad range of temperatures. In other words, under the appropriate scattering conditions a fast, positively charged particle in the vicinity of the ⟨100⟩ axis of Cl⁻ ions is channeled in an axial channeling regime. Based on the symmetry of the resulting effective potential (Figs. 2a–2d), it is conveniently approximated by a function of the type

$$U(\rho)=D_0\left(e^{-2\alpha\bar{\rho}}-2e^{-\alpha\bar{\rho}}\right),$$

$$\bar{\rho}=\frac{\rho-\rho_0}{\rho},\quad \rho=\sqrt{x^2+y^2},\tag{9}$$

where the parameters of the potential (9) have the following values in the CsCl crystal: $D_0=9.8$ eV, $\alpha=0.838$, and $\rho_0=0.46$ Å.

It is important to note that this approximation, as shown by comparison with direct numerical computations, is accurate to 1% or better for values of the potential $<-4$ eV. We also observe that values of the parameters of Cl⁻ and Cs⁺ close to those of the parameters of these ions in the free state were used in the numerical calculations of expression (5) in Lenz–Jensen distribution functions. For this reason the characteristics of the field of axial channeling of positrons should be expected to improve when more precise values of the crystal ion parameters are used in the calculations.

## TWO-DIMENSIONAL RELATIVISTIC POSITRON SYSTEM

The positron wave function in the axial channeling regime can be written in the form

$$\psi(\mathbf{r})=\frac{1}{2\pi}\exp\left(\frac{i}{\hbar}p_z z\right)\Phi(\rho,\varphi),\quad \mathbf{r}=(z,\rho,\varphi)\tag{10}$$

subject to the normalization condition

$$\int\psi^{*}(\mathbf{r})\psi(\mathbf{r})\,d\mathbf{r}=\delta(p_z-p_z')\,\delta_{nn'}\delta_{mm'},$$

$$\varepsilon<0,\quad n,m=0,1,2,\dots,\tag{10′}$$

where $\delta(P_z-P_z')$ is the Dirac delta function, $\delta_{nn'}$ is the Kronecker delta, $\varepsilon$ is the energy, and $\Phi(\rho,\varphi)$ is the bound state wave function.

Substituting Eq. (10) into the three-dimensional Schrödinger equation written in cylindrical coordinates $(x,\rho,\varphi)$, we find

$$\left(\frac{\partial^2}{\partial\rho^2}+\frac{2}{\rho}\frac{\partial}{\partial\rho}+\frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2}\right)\Phi(\rho,\varphi)$$

$$+\frac{2\mu}{\hbar^2}[\varepsilon-U(\rho)]\Phi(\rho,\varphi)=0,\tag{11}$$

where $\mu$ is the relativistic mass of the positron.

Now, proceeding from the symmetry of the potential (9), we write the solution of Eq. (11) in the form

$$\Phi(\rho,\varphi)=\frac{1}{2\pi\rho}e^{im\varphi}\chi(\rho).\tag{12}$$

Substituting Eq. (12) into (11), we obtain the following equation for the radial wave function:

$$\frac{d^2\chi}{d\rho^2}+\frac{2\mu}{\hbar^2}\left(\varepsilon-\frac{\hbar^2m^2}{2\mu\rho^2}-U(\rho)\right)\chi=0.\tag{13}$$

TABLE I.

| | | $E=5$ MeV | | |
|---|---|---|---|---|
| $n,m$ | 0 | 1 | 2 | 3 |
| 0 | $\varepsilon_{00}=-7.986$ | $\varepsilon_{01}=-5.893$ | | |
| 1 | | | $\varepsilon_{12}=-5.611$ | |
| 2 | | | | |
| 3 | | | | |

*Note:* Energy values are given in eV; places in the tables with no entries indicate that states with such quantum numbers do not exist.

TABLE III.

| | | $E=20$ MeV | | |
|---|---|---|---|---|
| $n,m$ | 0 | 1 | 2 | 3 |
| 0 | $\varepsilon_{00}=-8.368$ | $\varepsilon_{01}=-7.873$ | $\varepsilon_{02}=-6.388$ | $\varepsilon_{03}=-3.918$ |
| 1 | $\varepsilon_{10}=-8.535$ | $\varepsilon_{11}=-7.980$ | $\varepsilon_{12}=-6.318$ | $\varepsilon_{13}=-3.552$ |
| 2 | | | | $\varepsilon_{23}=-6.915$ |
| 3 | | | | |

*Note:* Energy values are given in eV; places in the tables with no entries indicate that states with such quantum numbers do not exist.

The subsequent analysis of Eq. (13) is analogous to that in Ref. 14. Since within the channel the coordinate $\rho$ does not differ much from $\rho_0$ for small quantum numbers, it is useful to expand the centrifugal term in (13) in powers of the parameter $\bar{\rho}$:

$$\frac{\hbar^2 m^2}{2\mu\rho^2}=\frac{m^2}{\gamma_0^2}D_0(c_0+c_1e^{-\alpha\bar{\rho}}+c_2e^{-2\alpha\bar{\rho}})+O(\bar{\rho}^3), \tag{14}$$

where we have introduced the notation

$$c_0=1-\frac{3}{\alpha}+\frac{3}{\alpha^2},\quad c_1=\frac{4}{\alpha}-\frac{6}{\alpha^2},$$

$$c_2=-\frac{1}{\alpha}+\frac{3}{\alpha^2},\quad \gamma_0^2=2\mu\frac{D_0\rho_0^2}{\hbar^2}. \tag{15}$$

Substituting Eq. (14) into (13), we arrive at an exactly solvable quantum mechanical model,[14] for which the radial wave function has the form

$$\chi(\bar{\rho})=y^{s}e_{1}^{-\frac{1}{2}y}F_{1}(a,c,y),\quad y=\frac{2\gamma_0}{\alpha}e^{-\alpha\bar{\rho}},\quad s=\frac{\beta}{\alpha}, \tag{16}$$

where

$$\beta^2=\beta_0^2+m^2c_0,\quad \gamma_1^2=\gamma_0^2-\frac{1}{2}mc_1^2,\quad \gamma_2^2=\gamma_0^2+\frac{1}{2}m^2c_2,$$

$$\beta_0^2=-\frac{2\mu\varepsilon\rho_0^2}{\hbar^2}>0,\quad a=\frac{1}{2}\left(\frac{2\beta}{\alpha}+1\right)-\frac{\gamma_1^2}{\alpha\gamma_2^2}. \tag{17}$$

The eigenvalues obey the equation[14]

$$\varepsilon_{nm}=\frac{\hbar^2}{2\mu\rho_0^2}\left\{\gamma_0^2+\alpha\gamma_0\left(n+\frac{1}{2}\right)-2^2\left(n+\frac{1}{2}\right)^2\right.$$
$$\left.+m^2-\frac{3(\alpha-1)}{\alpha\gamma_0}\left(n+\frac{1}{2}\right)m^2\frac{9(\alpha-1)^2}{4\alpha^4\gamma_0^4}m^4\right\}, \tag{18}$$

where $n$ is the vibrational quantum number, and $m$ is a quantum number characterizing the rotational motion. Tables I–IV give several values of the energy spectrum of transverse positron motion as a function of the quantum numbers $n$ (down the columns) and $m$ (along the rows) for various total positron energies $E$.

## CONCLUSION

We have shown that positively charged relativistic particles in ionic crystals of the CsCl type in the vicinity of the $\langle 100\rangle$ axis of $\mathrm{Cl}^-$ ions can be channeled in an axial channeling regime. We have shown by numerical analysis that the channeling potential has annular symmetry in this case, is situated in regions far from the crystal axes, and is essentially independent of the temperature of the medium. The latter property possibly means that the contribution of elastic scattering processes (both coherent and incoherent) to broadening of the energy levels of transverse motion must be insignificant. In other words, other mechanisms such as: a) inelastic processes with the excitation of electrons inside the crystal ions; b) radiative transitions between levels of transverse motion; c) band broadening, are responsible for broadening of the spectrum in this case. It is well known from the literature that these mechanisms operate independently of one another.

It is essential to note that radiative transitions are important for particles having higher energies $E\geqslant 10$ GeV (Ref. 15). As to the band broadening of levels due to the periodicity of the effective potential, Bazylev and Golovizin[16] have shown that it does not directly shorten the particle lifetime at

TABLE II.

| | | $E=10$ MeV | | |
|---|---|---|---|---|
| $n,m$ | 0 | 1 | 2 | 3 |
| 0 | $\varepsilon_{00}=-8.097$ | $\varepsilon_{01}=-7.082$ | $\varepsilon_{02}=-4.050$ | |
| 1 | | $\varepsilon_{11}=-9.330$ | $\varepsilon_{12}=-5.800$ | |
| 2 | | | | $\varepsilon_{23}=-8.322$ |
| 3 | | | | |

*Note:* Energy values are given in eV; places in the tables with no entries indicate that states with such quantum numbers do not exist.

TABLE IV.

| | | $E=30$ MeV | | |
|---|---|---|---|---|
| $n,m$ | 0 | 1 | 2 | 3 |
| 0 | $\varepsilon_{00}=-8.540$ | $\varepsilon_{01}=-8.213$ | $\varepsilon_{02}=-7.233$ | $\varepsilon_{03}=-5.602$ |
| 1 | $\varepsilon_{10}=-8.117$ | $\varepsilon_{11}=-7.758$ | $\varepsilon_{12}=-6.682$ | $\varepsilon_{13}=-4.889$ |
| 2 | | $\varepsilon_{21}=-9.789$ | $\varepsilon_{22}=-8.616$ | $\varepsilon_{23}=-6.662$ |
| 3 | | | | |

*Note:* Energy values are given in eV; places in the tables with no entries indicate that states with such quantum numbers do not exist.

a given level, and this phenomenon generally occurs for levels having high quantum numbers $n$ and $m$. It follows from these considerations that the most important mechanism of broadening of the spectral levels in the given problem, or the strongest dechanneling factor, is inelastic scattering involving excitation and ionization of ions of the strings..

It is reasonable to expect, therefore, that a) relatively low-energy ($E = 5-30$ MeV) positrons under the above-stated conditions will be superchanneled, thus forming stable two-dimensional relativistic positron systems; b) the proposed method can be used to generate positron systems from virtually all beam positrons in a crystal, a feat that is fundamentally impossible by the mixing of electron and positron beams in vacuum[7,8]; c) under the conditions of the given problem the parameters of the positron system can be varied considerably, including the creation of conditions for the resonance annihilation of a positron with an environmental electron, by means of external influences; d) the investigation of axial channeling of heavy positive ions can also be viewed as a promising undertaking.

The authors are grateful to members of the theoretical seminar at the Institute of Applied Problems in Physics of the National Academy of Sciences of Armenia for a fruitful discussion.

[1] J. A. Davies, J. Friesen, and J. D. McIntyre, Can. J. Chem. **38**, 1526 (1960).
[2] M. T. Robinson and O. S. Oen, Appl. Phys. Lett. **2**, 30 (1963).
[3] J. Lindhard, K. Dan. Vidensk. Selsk. Mat. **34**, No. 14 (1965).
[4] M. A. Kumakhov, Phys. Lett. **57**, 17 (1976).
[5] V. V. Beloshitsky and F. F. Komarov, Phys. Rep. **93**, 117 (1982).
[6] V. A. Bazylev and N. K. Zhevago, Usp. Fiz. Nauk **137**, 605 (1982) [Sov. Phys. Usp. **25**, 565 (1982)].
[7] M. Bertolotti and C. Sibilia, Lett. Nuovo Cimento **27**, 261 (1978).
[8] M. Bertolotti and C. Sibilia, Appl. Phys. **19**, 127 (1979).
[9] G. Kurizki and A. Friedman, Phys. Rev. A **18**, 512 (1988).
[10] A. S. Gevorkyan *et al.*, Zh. Tekh. Fiz. **59**(3), 54 (1989) [Sov. Phys. Tech. Phys. **34**, 285 (1989)].
[11] A. S. Gevorkyan *et al.*, ErFI Preprint No. ErFI-1051(14)-88 [in Russian], Erevan Physics Institute, Erevan (1988), 10 pp.
[12] A. S. Gevorkyan *et al.*, in *Proceedings of the Third All-Union Conference on the Investigation of Relativistic Particles in Crystals* [in Russian], Nalchik, (1988), p. 2.
[13] P. Gambos, *Statistical Theory and Its Application* [Russian translation] (IL, Moscow, 1951), 358 pp.
[14] S. Flügge, *Practical Quantum Mechanics*, (Springer-Verlag, Berlin, 1971) [Russian trans., Mir, Moscow (1974), 341 pp.].
[15] M. A. Kumakhov and R. Wedell, Phys. Status Solidi B **92**, 65 (1979).
[16] V. A. Bazylev and V. V. Golovizin, Zh. Éksp. Teor. Fiz. **82**, 1204 (1982) [Sov. Phys. JETP **55**, 700 (1982)].

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Scanning of a laser beam and purification of materials through the use of light-induced particle drift in semiconductors

N. N. Krupa and A. N. Pogorelyĭ

*Institute of Magnetism, Academy of Sciences of Ukraine, 252680 Kiev, Ukraine*

(Submitted October 14, 1996)
Zh. Tekh. Fiz. **68**, 121–124 (April 1998)

The light-induced drift of electrons, light-absorbing impurities, and defects in II–VI semiconductors is investigated experimentally, along with some potential practical applications of the phenomenon. It is shown that the light-induced drift of electrons induces a very pronounced change in the refractive index, $|\Delta n| \sim 0.01$, and can be used to implement effective scanning of nanosecond and picosecond laser pulses through frustration of total internal reflection. The light-induced drift of absorbing particles increases their density in the surface layer of the crystals, and this effect can be exploited in semiconductor technology. © *1998 American Institute of Physics.* [S1063-7842(98)02204-1]

## INTRODUCTION

The idea of making practical use of the phenomenon of variation of the refractive index of transparent media in the field of a laser beam was first propounded some time ago[1] and has been tested repeatedly since then in various engineering techniques to control the space–time characteristics of high-power light beams.

The most promising materials for the construction of high-speed nonlinear elements are semiconductors, because they are characterized by strong electron–phonon interaction with short relaxation times. The absolute value of the nonlinear variation of the refractive index $n$ should exhibit resonance growth in the vicinity of such electron transitions. The formation of a transverse gradient of the nonlinear variation of $n$ in semiconductors in the active zone of high-power laser beams has provided the means to achieve scanning,[2] modification of beam divergence by nonlinear lenses having a variable focal length,[3] and reduction of the duration[4] of nanosecond and picosecond laser beams when part of the beam is cut off by a diaphragm.

However, the practical application of developments of this kind is hindered by the fact that self-defocusing limits the magnitude of the effect for a negative variation of $n$, while self-focusing leads to damage of the material for a positive $\Delta n$.

Our aim in the present article is to propose for engineering application high-speed deflectors operating on the basis of frustration of total internal reflection,[5] along with a technological scheme for the purification of optical materials.[6] These developments have the distinguishing feature that they are based on the phenomenon of light-induced drift of nonequilibrium electrons and absorbing impurities in semiconductors. The light-induced drift of particles in the field of a high-power laser beam, which was predicted theoretically in Refs. 7 and 8, has been investigated experimentally for the most part in gaseous mixtures of atoms and molecules.

In addition, a well-known phenomenon in semiconductors is the entrainment of carriers by laser beams,[9,10] which can also apply to light-induced particle drift. The entrainment of electrons should raise their concentration in the region where the laser beams exit from the samples. Also, since the production of a high density of nonequilibrium electrons $N$ is the basic mechanism of negative variation of the refractive index of semiconductors under the influence of high-power laser pulses,[11]

$$\Delta n = -\frac{4\pi e^2 N}{n_0 m_e^x \omega^2}, \tag{1}$$

the light-induced drift of electrons should produce a significant drop in the refractive index and influence the total internal reflection of laser beams in semiconductors. Here $m_e^x$ is the effective mass of the electrons.

## EXPERIMENTAL PROCEDURE AND RESULTS

*Scanning of laser radiation by deflectors utilizing frustration of total internal reflection.* Nanosecond and picosecond pulses emitted by a ruby laser and a neodymium laser were used to investigate the total internal reflection phenomenon in CdSSe, ZnSe, and SiC crystals. The samples were cut in the form of triangular prisms in such a way that a laser beam normally incident on the input face would undergo total internal reflection at the opposite face and exit from the crystal (Fig. 1a). Measurements have shown that in samples characterized by two-photon absorption ($E_g < 2h\omega$) and in samples exhibiting fairly high impurity absorption or free-carrier absorption ($\alpha > 10\ \text{cm}^{-1}$) high-intensity radiation $I > 10\text{MW/cm}^2$, begins to propagate along the total-internal-reflecting face. Frustration of total internal reflection is observed not only in II–VI semiconductors, where the bulk excitation of nonequilibrium conduction electrons produces a decrease in the refractive index,[11] but also in SiC crystals, where nanosecond and picosecond laser beams are usually self-focused in the bulk[12] as a result of the nonlinearity polarizability of the conduction electrons.

These results indicate that the entrainment, or what might be called light-induced drift, of electrons in semicon-

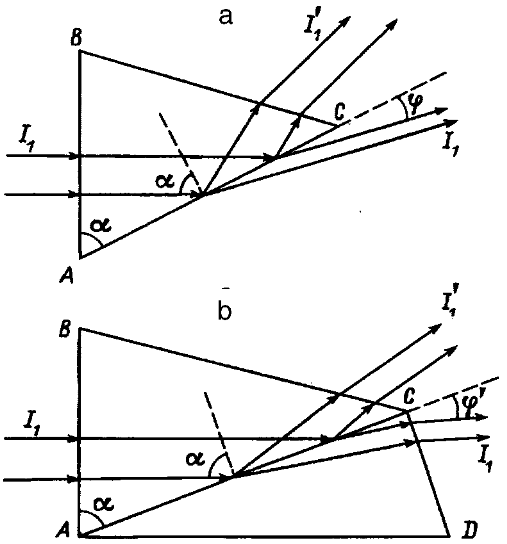

FIG. 1. Optical diagram of laser-beam deflectors operating on the principle of frustration of total internal reflection.

ductors produces a significant drop in the refractive index on the total-internal-reflecting surface. The magnitude of this decrement of the index, calculated from the angle through which the sample must be rotated to achieve total internal reflection in high-power laser beams, becomes very large: $\Delta n = 0.01 - 0.1$.

After laser beams have crossed the total-internal-reflecting face of the two-photon absorbing CdSSe and ZnSe crystals, they undergo continuous angular deflection. The maximum scanning angle, which can be estimated for Fig. 1a from the expression

$$\varphi_m = \arccos(1 - |\Delta n|/n), \tag{2}$$

has attained 15−17° for nanosecond pulses from ruby and neodymium lasers and 10° for picosecond pulses.

The scanning of laser beams after exiting from the total-internal-reflecting face is not observed in doped SiC semiconductors, where the frustration of total internal reflection is attributable to the light-induced drift of equilibrium conduction electrons. Self-focusing of the beam takes place, but the position of the axis of the radiation pattern remains essentially fixed within the duration of the pulse. Similar results have been obtained for highly doped II–VI crystals.

To increase the scanning angle, we have proposed a dual-prism deflector configuration[5] for nanosecond and picosecond laser pulses, operating on the principle of frustration of total internal reflection by the light-induced drift of nonequilibrium electrons in semiconductors (Fig. 1b). The first prism ABC is fabricated from a two-photon absorbing ($E_g < 2\hbar\omega$), high-resistance crystal (CdS or ZnSe for a ruby laser, CdSe for a neodymium laser). The face AC of this prism (the total-internal-reflecting face, $\sin\alpha = n_1/n_2$) is coupled through close optical contact to the face of the second prism ACD, which is made of a nonabsorbing material having a high refractive index (heavy grades of glass or, say, cubic zinc selenide). The angle ACD of this prism is close to 90°. In this deflector configuration the scanning angle is given by the expression

$$\varphi_m = \arcsin\{n_2 \sin[\arccos(1 - |\Delta n| n_2/n_1)]\}, \tag{3}$$

where $n_1$ and $n_2$ are the refractive indices of the first and second prisms.

Using cubic ZnSe and ZnC crystals, we have achieved almost 30° scanning of a ruby laser beam for nanosecond pulses and up to 20° for picosecond pulses.

It should be noted that in fabricating a scanning prism from a uniaxial, birefringent material, the laser beam must be directed into the prism along the optic axis, or at least linearly polarized light $E \perp C$ must be used.

The advantage of deflectors operating on the principle of frustration of total internal reflection, apart from their capability of achieving substantial scanning angles for nanosecond and picosecond laser pulses, is their superior operation with high-power, multimode laser beams. For sufficiently large unidirectional deflection angles there is essentially no significant spreading of the beam in the perpendicular direction.

*Light-induced drift of absorbing impurities and purification of semiconductors.* After type II–VI crystals had been irradiated with high-power nanosecond ruby and neodymium laser pulses, we observed a difference in the low-temperature luminescence spectra from the input and output surfaces of the irradiated samples. A similar difference in the rate of formation of luminescence centers on the input and output surfaces under the influence of two-photon absorbed laser pulses on II–VI semiconductors has been encountered in other papers,[12,13] but the authors, as a rule, dismiss the effect or mention the influence of the difference in the interference conditions on these surfaces,[14] which raises the intensity of the laser radiation on the exit surface by a factor $4n^2/(n+1)^2$.

To eliminate this influence, we have carried out investigations with a 100-W cw $CO_2$ laser. The low laser photon energy ($\hbar\omega_p \ll E_g$) lowers the probability of the direct formation of defects in the crystal lattice at low radiation intensities $I < 1000 \mathrm{W/cm^2}$. Nonetheless, after weakly absorbing CdSSe and ZnSe ($\alpha < 0.5\ \mathrm{cm^{-1}}$) crystals have been irradiated by a $CO_2$ laser for about 10 min, differences are observed in the recombination luminescence spectra from the input and output surfaces (Fig. 2).

At the input surface the half-width of the emission bands of the first phonon replica of a free exciton ($I_1$) and an exciton bound at a donor center ($I_2$) decreases somewhat, the structure of the luminescence of donor–acceptor pairs is more conspicuous, and a new band $Q$ appears, which, according to published data, corresponds to transitions in complexes with intrinsic defects or various impurity centers.[15]

At the output surface the luminescence of the first phonon replica of a free exciton ($I_1$) decreases, and the band itself, like the bound-exciton band, broadens considerably. Moreover, a new luminescence band ($I_7$) appears, which is usually attributed to in-center transitions in a quasi-isolated Cd atom (in CdS) or Zn atom (in ZnSe).[16]

The above-described asymmetry of the variation of the luminescence spectra of semiconductors after being irradiated by a high-power cw $CO_2$ laser indicates that under the influence of laser radiation the density of acceptor centers (of

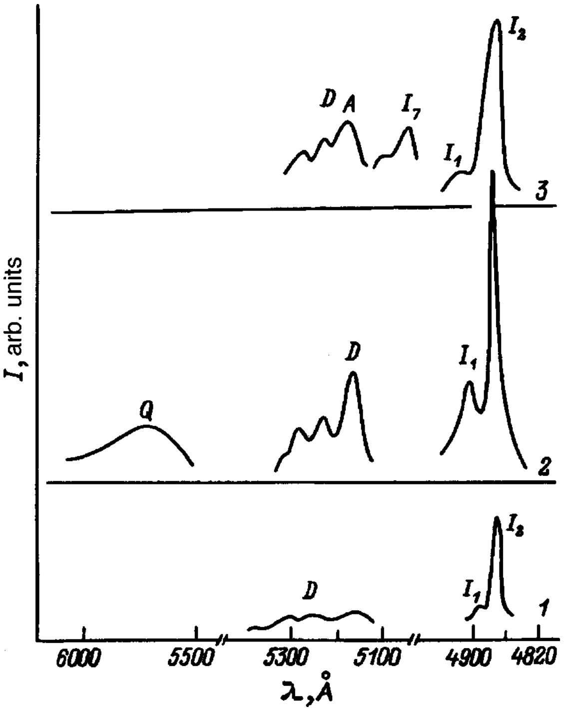

FIG. 2. Emission spectra of a CdS crystal at a temperature of 4.2 K. *1*) Before irradiation by a $CO_2$ laser; *2*) luminescence from the input surface after irradiation of the crystal by a $CO_2$ laser; *3*) from the exit surface after irradiation by a $CO_2$ laser; *D* denotes donor–acceptor pairs.

the interstitial sulfur or selenium type) increases on the input surface of the sample, and the density of donor centers (interstitial cadmium, zinc, or sodium-type impurity) increases on the output surface.

In addition to the changes in the luminescence spectra, a dark spot appears on the output surface in a large group of zinc selenide samples having a light-brown color and absorbing at wavelengths of 10.6$\mu$m and 0.1 cm after irradiation by a $CO_2$ laser. The outward appearance of the spot suggests that it is made up of crowded dark specks, whose size and density decrease from the center toward the periphery. At first the specks appear at the center of the exiting laser beam, and then with continued irradiation the area of the spot increases until it reaches the outer edge of the beam. The spot appears only on the output surface and, irrespective of the shape or finish of the surface (as-cleaved or polished), mimics the shape of the beam cross section. Mass-spectrometer measurements show that the carbon concentration on the surfaces of these samples in the exit zone of the laser beam increases at least by an order of magnitude after laser irradiation. The measured temperature difference between the input and output surfaces of the crystals never exceeds a half a degree during irradiation.

These results show that high-intensity radiation from a $CO_2$ laser imparts a directional motion to the light-induced drift of absorbing impurities and intrinsic defects in II–VI semiconductors. Drift can be imparted to charged defects by the tractor field of an electron swarm displaced along the laser beam by the entrainment effect. Drift is imparted to neutral particles by the difference in the cross section of interaction of such a defect with the semiconductor lattice in the ground and excited states and can occur when the frequency of the laser beam does not coincide exactly with the frequency of transition of the defect from the ground state to the excited state or when the transition band is highly asymmetric within the limits of the laser emission line.

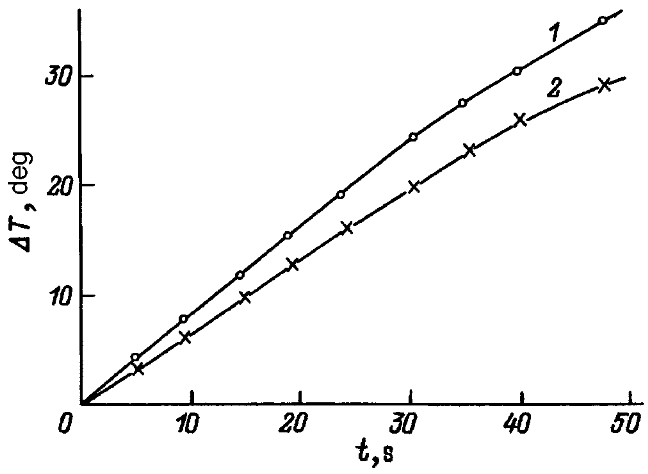

FIG. 3. Temperature increment in heating of a ZnSe crystal by a $CO_2$ laser beam (100 W) versus time. *1*) Before treatment of the crystal; *2*) after two purification operations.

A technological scheme for the purification of zinc selenide has been proposed on the basis of the above-described effect and is widely used in the fabrication of optical elements for high-power infrared lasers. The scheme essentially entails the following. Zinc selenide crystals are polished mechanically and chemically, and the light zone is irradiated with a high-power cw laser for several tens of minutes. The input and output surfaces of the sample are then polished again both mechanically and chemically. The absorption coefficient of the treated crystal can be lowered by repeating this entire operation several times. During laser irradiation the sample can be placed in an inert gas atmosphere, permitting the laser power to be increased and thereby enhancing the purification efficiency. The distinguishing feature of this purification technique is that it provides the best means for eliminating absorbing impurities and defects from the interior volume.

Figure 3 shows the decrease in the heating rate of a ZnSe sample in the calorimetric chamber after two purification runs, graphically illustrating the reduction in the absorption coefficient of the sample.

[1] P. D. McWane, Nature **211**, 1081 (1966).
[2] A. A. Borshch, M. S. Brodin, V. I. Volkov, and V. V. Ovchar, Kvantovaya Élektron. **2**, 602 (1975) [Sov. J. Quantum Electron. **5**, 340 (1975)].
[3] A. A. Borshch, M. S. Brodin, V. I. Volkov, and N. N. Krupa, Zh. Tekh. Fiz. **48**, 2273 (1978) [Sov. Phys. Tech. Phys. **23**, 1356 (1978)].
[4] N. N. Krupa, N. Ya. Nedbaev, P. A. Petrenko, and A. A. Statsenko, USSR Author's Certificate No. 10901145, International Patent Classification (IPC) 3 G 02 1/37, application dated June 18, 1982, published January 3, 1984.
[5] V. T. Bagatskiĭ, N. N. Krupa, and V. I. Lomakin, USSR Author's Certificate No. 1259844, International Patent Classification (IPC) G 02 F 1/09, application dated August 16, 1984, published July 29, 1985.
[6] N. N. Krupa and V. I. Lomakin, USSR Author's Certificate No. 1294036, International Patent Classification (IPC) C 03 B 33/29/45, application dated February 13, 1985, published July 29, 1985.
[7] F. Kh. Gel'mukhanov and A. M. Shalagin, JETP Lett. **29**, 711 (1979).

[8] F. Kh. Gel'mukhanov and A. I. Parkhomenko, Inst. Avtomat. Élektrometr. SO AN SSSR Preprint No. 340 [in Russian] (Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences, 1987).

[9] N. A. Bryzhnykh, A. A. Grinberg, and É. Z. Imanov, Fiz. Tverd. Tela (Leningrad) **5**, 1735 (1963) [Sov. Phys. Solid State **5**, 1261 (1963)].

[10] A. M. Danishevskiĭ, A. L. Kosmal'skiĭ, and S. M. Ryvkin, Zh. Éksp. Teor. Fiz. **58**, 544 (1970) [Sov. Phys. JETP **31**, 292 (1971)].

[11] A. A. Borshch, M. S. Brodin, and N. N. Krupa, Zh. Éksp. Teor. Fiz. **70**, 1805 (1976) [Sov. Phys. JETP **43**, 940 (1976)].

[12] A. A. Borshch, M. S. Brodin, and V. I. Volkov, Zh. Éksp. Teor. Fiz. **72**, 938 (1977) [Sov. Phys. JETP **45**, 490 (1977)].

[13] M. S. Brodin, N. A. Davydova, and I. Yu. Shabliĭ, Izv. Akad. Nauk SSSR, Ser. Fiz. **46**, 1037 (1982).

[14] N. Boling, P. Braunlich, A. Schmid, and P. Kelly, Appl. Phys. Lett. **27**, 191 (1975).

[15] V. V. Dyakin and N. S. Khalilova, Ukr. Fiz. Zh. **26**, 1297 (1981).

[16] Yu. M. Émirov, S. S. Ostapenko, M. A. Rizakhanov, and M. K. Sheĭkman, Fiz. Tekh. Poluprovodn. **16**, 1371 (1982) [Sov. Phys. Semicond. **16**, 879 (1982)].

Translated by James S. Wood

# Electron-beam tomography of the density of a gas in hypersonic flow around objects

A. V. Likhachev, A. A. Maslov, S. G. Mironov, and V. V. Pikalov

*Institute of Theoretical and Applied Mechanics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia*

(Submitted December 23, 1996)
Zh. Tekh. Fiz. **68**, 125–133 (April 1998)

A procedure is developed for the tomographic reconstruction of the distribution of the gas density from measurements data on the attenuation of a sensing beam of fast electrons. The measurements and reconstruction are carried out for symmetrical and asymmetrical conical objects in a low-density hypersonic flow ($M = 21$). *© 1998 American Institute of Physics.* [S1063-7842(98)02304-6]

## INTRODUCTION

Hypersonic flow around objects at moderate Reynolds number in the presence of strong viscid-inviscid interaction has important bearing on the development of heat-shielding systems for returnable space vehicles and multimission space systems. Of special interest are flows involving the interaction of two or more objects. An important parameter is the gas density, whose measurement in rarefied hypersonic flow poses a complex methodological problem. On the other hand, experimental data on the spatial distribution of the density provide a foundation for improving numerical methods in hypersonic aerodynamics. These considerations call for the development of procedures by which to measure gas density near objects in complex three-dimensional flow environments.

The probing of gas flows by beams of charged particles one method of direct, nondisruptive density measurement. The application of electron-beam fluorescence,[1] which is widely known in rarefied gas dynamics, is justified only at low densities $n < 10^{21}\ \mathrm{m}^{-3}$. Far higher flow densities are achieved in existing hypersonic wind tunnels, where the influence of secondary electrons and intermolecular collisions is strong. This consideration limits the method to two-dimensional flows and offsets the localized character of the measurements.[2] The density measurement problem is solved in part by using x-ray bremsstrahlung from electrons instead of optical radiation,[3] but the measurement time is prolonged considerably by the low intensity of the recorded radiation. In addition, difficulties persist in plotting a calibration curve for three-dimensional gaseous object in dense wind tunnel flow.

An alternative method is electron-beam densitometry, which is based on the attenuation of a slender beam of fast electrons in a gas during elastic and inelastic scattering by molecules. The method is not very sensitive to the composition of the gas, it does not require complex recording instruments, and the measurements can be performed at high speed. It has been employed in physical measurements since the end of the fifties.[4] It has been used to investigate one-dimensional,[5] two-dimensional,[6] and axisymmetrical[7,8] low-density flows. The main drawback of the method is the tendency of the gas density information to be integrated along the length of the diagnostic beam. The latter problem is solved by applying the tomographic approach to the problem of reconstructing local density. For example, Ivanov[7] has used the Abel transform to reconstruct the density field of a gas in the vicinity of a sphere. Multiaspect transmission tomography methods have now been fairly well developed; they are used extensively in gas and plasma diagnostics[9] and can be adapted to the conditions of electron-beam probing of a gas flowing around an object. Problems arise from the methodological viewpoint in connection with the presence of zones shadowed by the model and the finite dimensions of the beam and the electron detector.

In this paper we describe an experimental procedure for the multiaspect probing of a rarefied hypersonic flow and give a tomographic reconstruction algorithm, along with the results of reconstructing a two-dimensional gas-density field.

## 1. SETUP OF THE TOMOGRAPHIC EXPERIMENT

The measurements were carried out in the I-327 hypersonic wind tunnel at the Institute of Theoretical and Applied Mechanics, Siberian Branch of the Russian Academy of Sciences (ITPM SO RAN).[10] The flow parameters were as follows: stagnation temperature $T_0 = 1100$ K, stagnation pressure $P_0 = 8$ MPa, freestream Mach number $M = 21$, unit Reynolds number $\mathrm{Re}_1 = 6 \times 10^5\ \mathrm{m}^{-1}$. Two models were used in the measurement: a blunted cone and a blunted semielliptical cone, both made of aluminum. The blunting radius of the cones was 1.5 mm, and the models had a length of 0.1 m. The radius of the base of the regular cone was 13.4 mm, and the vertex angle was 7°. The base of the semielliptical cone was a semicircle of radius 1.34 mm joined to a semiellipse with principal axes of length 13.4 mm and 6.7 mm. The measurements were performed in the plane situated at 0.065 m from the nose of each model.

The measurement arrangement is shown in Fig. 1. An electron beam of energy 20 keV and current intensity 0.1 mA is split in two (object beam *1* and reference beam *2*) by feeding an alternating square-wave voltage of the meander type with a frequency of 215 Hz to the magnetic system used to control the position of the beam. During the measurements

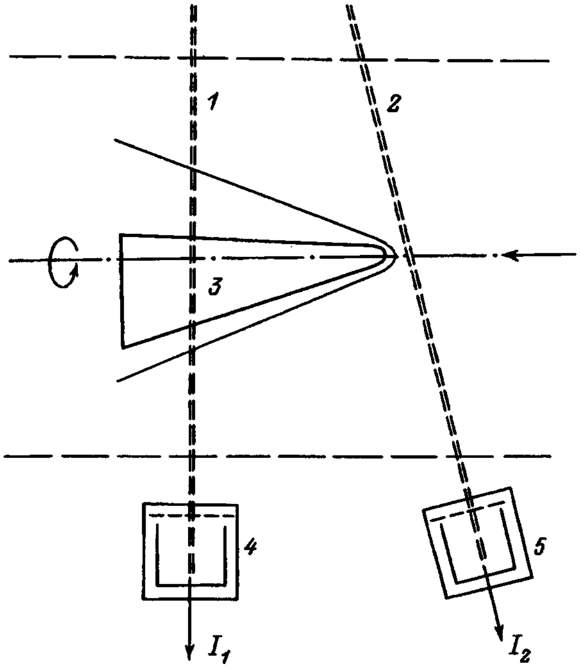

FIG. 1. Experimental arrangement for the tomographic investigation of hypersonic flow around a model.

beam 1 passes through the flow region near the model *3*, and beam *2* interacts only with the undisturbed flow in front of the model. The formation of two beams from a single beam prevents the measurement accuracy from being influenced by beam-current and flow-density fluctuations and permits the scattering of the beam in the incident flow to be taken into account. The electron current of each beam is recorded by the collectors *4* and *5*, which are set up outside the flow and have a circular diaphragm of diameter 4.5 mm. Each collector has a grid carrying a voltage of $-20$ V to delay secondary electrons. Scanning of the flow field by beam 1 is implemented by slewing the model across the beam between the limits of $-15$ mm and $+15$ mm relative to the flow axis. Scanning of the flow field with respect to angle of observation (aspect) is achieved by rotating the model about its own axis at a rate of 1 turn/s. The data on the displacement of the model across the flow and the angular position of the model are recorded from the linear and circular rheochords and are recorded simultaneously with the collector currents on an NO67 multichannel magnetograph.

The primary processing of the measurement data entailed the formation of data arrays, including the impact parameter $p$, the angle of rotation of the model $\phi$, and the object-beam and reference-beam collector currents $I_1$ and $I_2$ with uniform time division of the readings. The rms currents at the modulation frequency were determined accordingly. This approach was used to avoid stray currents and background biasing of the amplifier and magnetograph output voltages.

The dependence of the attenuation of the electron-beam current on the density of the gas and the scattering path length was determined experimentally with the gas at rest. For this purpose the beam current was measured during variation of the distance $l$ from the beam-entry aperture of the working chamber to the electron collector and during variation of the nitrogen density $n$ in the working chamber of the apparatus. The measurements were carried out in the range of the parameter $nl=(3.6-12)\times 10^{20}$ m$^{-2}$, spanning all the freestream values of this parameter. The measurements showed that the dependence of the electron current on $n$ and $l$ for the chosen diameter of the entrance to the collector is well characterized by the exponential law

$$I=I_0\exp(-\sigma nl). \tag{1}$$

The effective scattering cross section $\sigma$ determined from the measurement data is $\sigma=(1.9\pm 0.05)\times 10^{-21}$ m$^2$ for 20-keV electrons. The freestream gas density is estimated from the scattering cross section and the diameter of the flow. It is equal to $5.6\times 10^{21}$ m$^{-3}$, which is 18% lower than the value determined from the isentropic relation for the given Mach number. The discrepancy can be attributed to the somewhat higher static temperature established in the undisturbed flow.

The sensitivity and resolving power of the method depend on the choice of diameter $d$ of the collector diaphragm. For a beam with a Gaussian radial current distribution the maximum sensitivity to beam broadening is attained when the relation $d\simeq\sqrt{2}\gamma$ holds. Here $\gamma$ is a characteristic radius of the beam at the entrance to the collector; it depends on the electron energy, the gas flow density, and the length of the scattering zone.[11] The value of $\gamma$ has been determined by measuring the attenuation of the beam current in the freestream flow as the diameter of the collector diaphragm was varied from 1 mm to 10 mm. The diaphragm was then set for the sensitivity-optimum aperture diameter, which was 4.5 mm in the given experiments.

The measurements on the circular cone served as a test problem for fine-tuning the algorithm for reconstruction of the density field. The data from measurements of the beam current attenuation in transition from the model shadow zone into the penumbra were used to determine the radial profile of the sensing beam in the vicinity of the model. The results of reconstructing the beam profile are shown in Fig. 2b (curve *1*). It is evident from the figure that the beam is narrow, and its width at half-maximum does not exceed 0.5 mm. The slight asymmetry of the distribution is possibly associated with the presence of fast electrons in the beam, some of which enter into tangential interaction with the surface of the metal model.

## 2. DERIVATION OF AN EQUATION RELATING THE VARIATION OF THE GAS DENSITY TO THE MEASURED CURRENTS, ASSUMING A FINITE COLLECTOR APERTURE

If we adopted the same mathematical model as in Ref. 7 to relate the density of the gas along the axis of the electron beam to the recorded current, the experimental data would be related by the equation

$$\frac{I_1}{I_2}=\exp\left(-\sigma\int_{-\infty}^{\infty}\Delta n(s)ds\right). \tag{2}$$

In Eq. (2) $\sigma$ is the electron scattering cross section, and $\Delta n(s)$ is the change in the gas density along the axis of the electron beam due to the intrusion of the aerodynamic

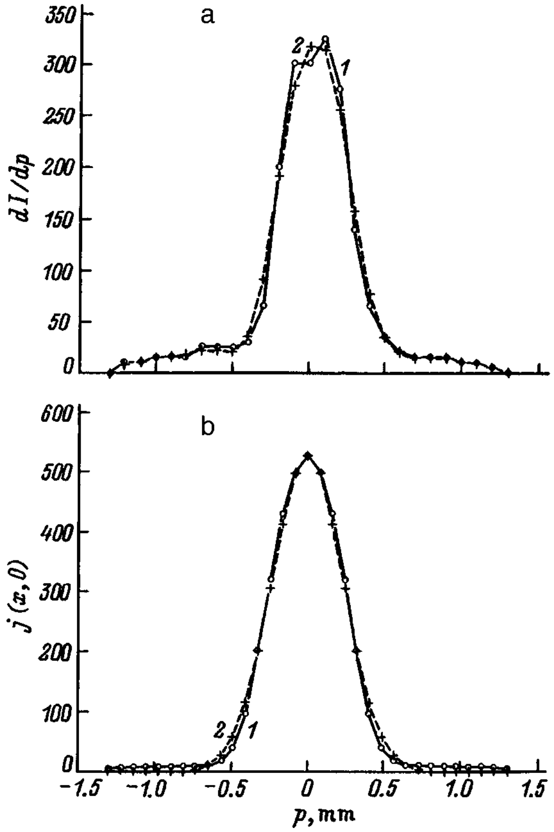

FIG. 2. a: Projection of the current density distribution in the beam: *1*) initial curve; *2*) spline-smoothed curve, assuming type 2 noise at the 3% level. b: *1*) Cross section of the reconstructed current density distribution at the $y$ axis; *2*) Gaussian function approximating the distribution.

model; the currents $I_1$ and $I_2$ are shown in Fig. 1. Equation (2) has been derived in the approximation of an infinitely small detector aperture. However, the relatively large size of the collector diaphragm casts doubt on such an approximation. We have investigated another mathematical model that conforms more closely to the experimental conditions. Since electrons are scattered mainly at small angles in the given situation, undergoing approximately one collision in transit from the source to the collector, we disregard the variation of the projection of the electron velocity onto the beam axis, thereby assuming, in effect, that its current through cross sections perpendicular to the axis remains constant. We denote the distribution of the component of the beam current density parallel to the beam axis by $j(r,\varphi)$ and, for brevity, refer to it below as the current density distribution in the beam or simply the beam current density. If the collector detects all electrons incident on a circular diaphragm of radius $d$ situated in the plane perpendicular to the beam axis, the current on it is

$$I=\int_0^{2\pi}\int_0^d j(r,\varphi)r\,dr\,d\varphi. \tag{3}$$

We have assumed that the scattering of electrons is a stochastic process and, accordingly, that the beam current density has a gaussian profile. The results of our experimental determination of the beam profile in the next section corroborate this assumption. Denoting the half-width of the Gaussian distribution by $\gamma(s)$ (bearing in mind that $\gamma$ varies along the beam axis) and normalizing the beam current to the current at the source $I_0$, we write the distribution of the beam current density in the form

$$j(r,\varphi)=\frac{I_0}{\pi\gamma^2(s)}\exp\left(-\frac{r^2}{\gamma^2(s)}\right). \tag{4}$$

After the beam profile (4) is substituted into Eq. (3), the integration over $r$ is elementary and gives

$$I=I_0\left[1-\exp\left(-\frac{d^2}{\gamma^2(s)}\right)\right]. \tag{5}$$

The current density on the beam axis at the entrance to the collector diaphragm $j(0,\varphi)$ and at the exit from the source $j_0$ are related by Eq. (2) with the appropriate replacement of the currents by their densities on the axis. We therefore have

$$j(0,\varphi)=\frac{I_0}{\pi\gamma^2(s)}=j_0\exp\left(-\sigma\int_{-\infty}^{\infty}n(s)ds\right). \tag{6}$$

Determining $\gamma^2(s)$ from Eq. (6) and substituting into (5), we have

$$I/I_0=1-\exp\left[-\frac{\pi d^2 j_0}{I_0}\exp\left(-\sigma\int_{-\infty}^{\infty}n(s)ds\right)\right]. \tag{7}$$

The two currents $I_2$ and $I_1$ corresponding to the beams transmitted through the undisturbed and disturbed flows were recorded in the experiment. The currents $I_2$ and $I_1$ obey Eq. (7) with the corresponding (to each) value of the density. We transform the two resulting equations and, dividing the first by the second, we obtain

$$\frac{\ln(1-I_1/I_0)}{\ln(1-I_2/I_0)}=\exp\left(-\sigma\int_{-\infty}^{\infty}\Delta n(s)ds\right). \tag{8}$$

If the numerator and denominator on the left side of Eq. (8) are expanded in series in $I_1/I_0$ and $I_2/I_0$, respectively, and only first-order terms are retained, we obtain Eq. (2). The legitimacy of discarding all other terms of the expansion for measurements performed in our experiment is confirmed by direct comparison of the left sides of Eqs. (2) and (8) for all the recorded currents $I_1$ and $I_2$. The current $I_0$ is taken as the average over several control measurements. The results of calculations show that the left sides of Eqs. (2) and (8) differ at most by 15% and, in the average over all measurements, by 6–7%. This accuracy is roughly consistent with the accuracy of the measurements themselves. In the given situation, therefore, the approximate equation (2) can be used to determine the gas density, and in fact we have done so in the present study.

## 3. RECONSTRUCTION OF THE ELECTRON BEAM PROFILE

An idea set forth in Ref. 12 has been elaborated to reconstruct the profile of the electron beam. Experimental data obtained for the symmetrical model in a region where it partially overlaps the beam are used here.[1)] We disregard the sloping of the generator of the cone, i.e., in the region of the beam we replace the cone by a cylinder. We let the beam

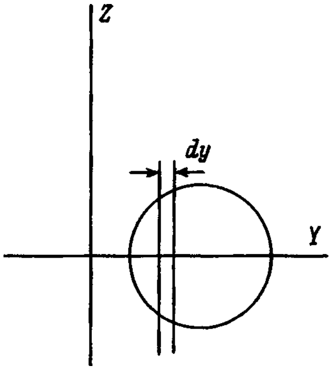

FIG. 3. Diagram illustrating the method of determination of the current density distribution in the cross section of the beam. The cross section of the beam is shown in the form of a circle.

move perpendicularly to the generator of the cone, along the $y$ axis, as shown in Fig. 3. In the figure the beam propagates from left to right, the part of the beam shadowed by the model is to the left, and the tomographic data are recorded perpendicularly to the plane of the figure, along the $x$ axis. We wish to relate the variation of the collector current for an infinitely small displacement of the beam to the variation of its cross section perpendicular to the model. In this case we can write

$$dI=\int_{-\infty}^{y+dy}\int_{-\infty}^{\infty}j(y,z)dydz-\int_{-\infty}^{y}\int_{-\infty}^{\infty}j(y,z)dydz$$
$$=\int_{y}^{y+dy}\int_{-\infty}^{\infty}j(y,z)dydz. \tag{9}$$

Dividing Eq. (9) by $dy$, we obtain

$$\frac{dI}{dy}=\int_{-\infty}^{\infty}j(y,z)dz, \tag{10}$$

i.e., the derivative of the object-beam current in the zone of partial overlap with the model is the projection (see, e.g., Ref. 9) of the current density distribution in the beam. This derivative is shown in Fig. 2a, curve *1*. The resulting projection of the current density distribution in the beam is smoothed by a cubic spline[13] on the assumption that it is distorted by random noise with a variance equal to 3% of the maximum (Fig. 2a, curve *2*). For the tomographic reconstruction of the current density distribution in the beam cross section near the model we assume that the distribution has circular symmetry. Reconstruction is executed separately for the left and right sides of curve *2* by means of the ART1 algorithm (see below). The results are then averaged.

The cross section of the reconstructed (assuming circular symmetry) current density distribution in the beam at the $Y$ axis is shown in Fig. 2b (curve *1*). Curve *2* in this figure is a Gaussian function approximating the distribution. The parameters of the Gaussian curve are determined by the least-squares method; in particular, its half-width is approximately equal to 0.275. The normalized standard deviation between curves *1* and *2* in Fig. 2b is 4.7%.

The current distribution density in the beam is used below to refine the reconstruction of the field of the density variation in a shock wave (Sec. 6).

## 4. TOMOGRAM RECONSTRUCTION ALGORITHM

In the experiments we have measured the currents of beams having various orientations relative to the model, and the axes of all the beam were in the same pane perpendicular to the axis of the model. We refer to this plane below as the cross-sectional plane. Tomographic methods can be used to reconstruct a two-dimensional scalar density-variation field from the experimental data; this field is naturally ascribed to the cross-sectional plane. However, the values so obtained will be averaged along the coordinate perpendicular to the cross-sectional plane over a scale of the order of the effective beam diameter. In discussing the density variation in the cross-sectional plane below, we have these averaged values specifically in mind.

From the mathematical standpoint the problem reduces to the reconstruction of a function of two variables from the set of its integrals along certain straight lines. A distinctive attribute of the stated problem associated with processing of the experimental data is the presence of an opaque model in the flow. The electron beam is absorbed when it enters the model, and the recorded current becomes equal to zero, whereas the integral of the density variation along the corresponding straight line has a nonzero value in general. Tomographic problems of this kind have been investigated previously.[9]

We introduce a coordinate frame rigidly attached to the model. The origin is located in the cross-sectional plane at the point where the axis of the model passes through it. We orient the system in such a way that the $X$ and $Y$ axes lie in the cross-sectional plane. We introduce the notation

$$\Delta n(x,y)=n_2(x,y,0)-n_1(x,y,0), \quad f_i=\ln\frac{I_{1i}}{I_{2i}}, \tag{11}$$

where $I_{1i}$ and $I_{2i}$ are the measurements corresponding to the $i$th beam.

We partition the zone of reconstruction of the function $\Delta n(x,y)$ into square cells, or pixels. The function $\Delta n(x,y)$ is assumed to have a constant value in each pixel. In this case the integral in (2) is replaced by a finite sum over the pixels through which the axis of the beam passes. This formulation reduces the tomographic reconstruction of the function $\Delta n(x,y)$ to the inversion of the system of linear algebraic equations

$$A\Delta n=f. \tag{12}$$

Here $A$ is an $I\times J$ matrix, $\Delta n\in R^J$ and $f\in R^I$are vectors corresponding to the reconstructed function and the experimental data, respectively, $J$ is the number of pixels, and $I$ is the number of sensing beams. The element $a_i^j$ of the matrix $A$ is defined as the length of the intersection of the axis of the $i$th beam with the $j$th pixel.[2)] We have chosen the iterative algorithm ART1 (Algebraic Reconstruction Technique) to invert the system (12), because, first, it has proved itself in reconstruction from insufficient data and, second, it can be

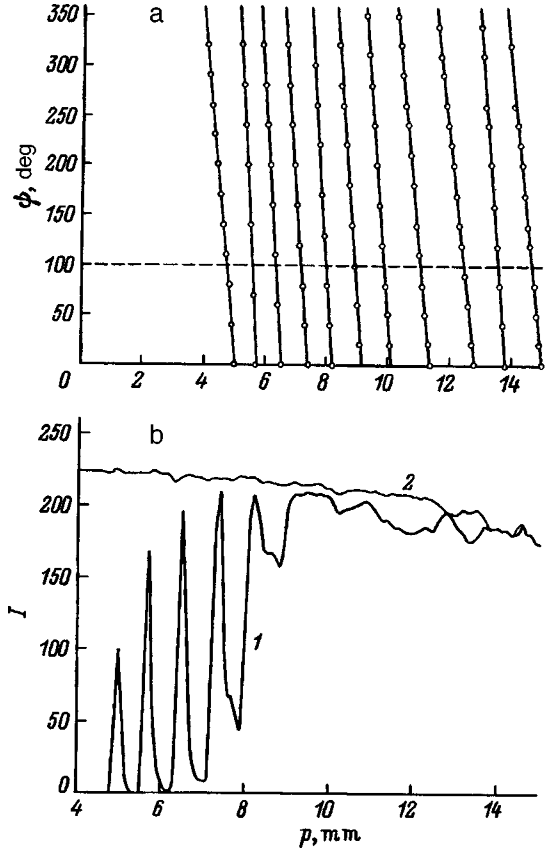

FIG. 4. a: Set of data in coordinates $p,\phi$, where $p$ is the impact parameter along the beam axis, and $\phi$ is the angle of rotation of the model; b: recorded currents versus impact parameter.

used without major revision in the presence of an opaque object (model). Following Ref. 14, we write the $(k+1)$st iterative solution for ART1 in the form

$$\Delta n^{(k+1)}=\Delta n^{(k)}+\lambda^{(k)}\frac{f_{i(k)}-(a^{i(k)},\quad \Delta n^{(k)})}{||a^{i(k)}||^2}a^{i(k)},$$

$$||a^{i(k)}||\neq 0,\quad \Delta n^{(k+1)}=\Delta n^{(k)},\quad ||a^{i(k)}||=0, \tag{13}$$

Here $a^{i(k)}$ is the $i(k)$th row of the matrix $A$, $\lambda^{(k)}$ is a relaxation parameter, and $i(k)=[k(\mathrm{mod}\ I)+1]$, i.e., the rows of the matrix $A$ are cyclically permuted. The scalar product and the norm in $R^J$ are defined in the usual way. It is stated in Ref. 14 that the iterative process (3) converges if $0<\lambda\leqslant 2$ for any initial approximation $\Delta n^{(0)}\in R^J$.

## 5. PRELIMINARY DATA PROCESSING

*a) Interpolation of the measurement data and construction of projections.* The results in this section pertain to the processing of data obtained for the semielliptical model. For the symmetrical model the measurements formed a single projection at once and did not need additional interpolation.

To describe the data set, we introduce a plane with coordinates $p,\phi$ (recall that $p$ is the impact parameter along the beam axis, and $\phi$ is the angle of rotation of the model). Each measurement in this plane is represented by a single point. Figure 4a shows the set of points corresponding to the first series of measurements is shown in the $p,\phi$ plane. Figure 4b shows the data from the first series of measurements. Curves *1* and *2* refer to points of the object and reference beams, respectively. The coordinate $p$ is plotted along the horizontal axis, and the measured current is plotted along the vertical axis. The angle varies periodically as a function of $p$ in this case; it can be estimated by comparing Figs. 4a and 4b.

In Fig. 4b the interval $p\approx 4-9$ mm corresponds to partial absorption of the object beam by the model. The peaks correspond to beams transmitted from the elliptical part of the model. The interval $p\approx 9-13$ corresponds to object beams cutting across the shock-wave region. At $p\geqslant 13.5$ the currents for the reference and object beams essentially coincide; this means that the disturbances generated by the model become insignificant in this range.

In numerous computer experiments[9,14] it has been shown that the ART1 algorithm yields a better quality of reconstruction when the latter is based on a set of projections with a uniform angular distribution, each projection being determined on a uniform grid.

It is evident from Fig. 4a that the data in $(p,\phi)$ space fit definite straight lines to within small discrepancies due to random errors. The slopes of these lines are determined from the coordinates of their corresponding points by the least-squares method. The experimental points are then projected onto the resulting lines. In $(p,\phi)$ space a projection corresponds to a line $\phi=\mathrm{const}$. The points of intersection of the line $\phi=\mathrm{const}$ with the best-fit lines of the experimental data are assigned values determined by linear interpolation along the lines (Fig. 4a). This procedure results in the construction of a projection at the angle $\phi=\mathrm{const}$ specified on a nonuniform grid. The transition to a uniform grid on the projection is also made by linear interpolation.

*b) Correction of projections in the penumbra.* Equation (2) needs to be refined near the boundary of a model that is opaque to the electron beam. Indeed, owing to the finite dimensions of the beam, part of it absorbed by the model, and part of it is incident on the collector. The recorded current decreases considerably in this case. Consequently, a penumbra zone is formed around the model, corresponding to the interval 4–9 mm in Fig. 4b. In the penumbra the drop in the current $I_1$ is caused mainly by model absorption, which is generally a much stronger effect than scattering by shock waves. As a result, the projection increases abruptly near the boundaries of the model. Tomographic reconstruction from such projections is found to be unsatisfactory. However, the large values on the projections in the penumbra cannot be discarded (e.g., by setting them equal to zero), because the shock waves come so close to the model that the majority of the electron beams partially absorbed by the model also pass through the shock waves.

The problem of isolating the effect associated with model absorption has been solved as follows in our study. The projections are modified by introducing the compensating increment

$$f(p)=\begin{cases} f(p)+\psi(p,p_b) & |p-p_b|\leqslant r_0,\\ f(p) & |p-p_b|>r_0. \end{cases} \tag{14}$$

In Eq. (14) $p_b$ is the coordinate of the boundary of the model on the projection closest to the point with coordinate $p$, and $r_0$ is the effective beam radius near the model. The function $\psi(p,p_b)$ is given by the equation

$$\psi(p,p_b)=\ln\left(\frac{\int_{S_2} j(r,\phi)ds}{\int_{S_1} j(r,\phi)ds}\right). \tag{15}$$

The distribution of the current density in the beam $j(r,\varphi)$ has been reconstructed from the experimental data (Sec. 3). From this result we estimate its effective radius to be $r_0 \approx 0.5$ mm. In Eq. (15) $S_1$ is the area of the disk of radius $r_0$, and $S_2$ is the area of the segment of this disk bounded by the generator of the cone corresponding to the model. Physically the function $\psi(p,p_b)$ can be interpreted as the logarithm of the ratio of the energy of the transmitted part of the beam to its initial energy.

## 6. RESULTS OF TOMOGRAPHIC RECONSTRUCTION AND DISCUSSION

A square centered at the origin with a side of 30 mm is chosen as the reconstruction zone. The reconstruction zone is partitioned into 65×65 square pixels. Each pixel therefore has a side ≈0.46 mm, which is of the order of the experimentally determined beam half-width (Sec. 3). For the semielliptical model the two-dimensional distribution of $\Delta n(x,y)$ is reconstructed separately from data obtained in three series of measurements and also from the data averaged over all three series. For each series of experimental data 18 projections distributed uniformly (every 10°) in the interval from 0° to 180° (it is obvious that projections at angles $\Theta$ and $\Theta+180°$ are equivalent) are generated for each series of experimental data by the interpolation described in Sec. 5a. Each projection is determined on a uniform 33-node computational grid by interpolation; the spacing of the nodes is ≈0.92 mm. The projection is then compensated by the procedure described in Sec. 5b. After each iteration the resulting distribution is smoothed in a sliding 3×3 window. The iterative process is terminated when the residual between the projections and the pseudoprojection no longer decreases.

Reconstruction based on different series of experimental data produces similar shock wave distribution patterns in the vicinity of the model. However, the tomograms exhibit amplitude peaks and troughs of a random nature, most likely associated with an insufficiency of data and errors in compensation of the magnitude of the projection. Averaging of the projections over all three series helps to eliminate artifacts of this kind. Figure 5 shows in axonometric projection the reconstructed variation of the density field in the cross-sectional plane in the averaged projections. These results and those that follow are given in the form of the ratio $(\Delta n + n_\infty)/n_\infty$, where $n_\infty$ is the freestream density (it is defined in Sec. 1).

It is evident from Fig. 5 that the density distribution matches the physical picture of the phenomenon. Clearly visible is the asymmetry of the density distribution due to the elliptical part of the model, in whose vicinity we observe a reduction of the compression shock and the formation of a smoother density profile along the normal to the surface. For all practical purposes, the region of rarefaction near the model, whose existence is inferred from theory and the results of measurements on similar objects (see, e.g., Ref. 15), is not reconstructed. These distortions are attributable primarily to shortcomings of the projection data. It is evident from Fig. 4a that in reality data have been recorded on the average with a step of 1.5–2 mm for each projection. Consequently, only one or two readings are associated with each projection onto the shock wave region, and such a figure is inadequate for good tomographic reconstruction. Another factor is the beam width, which is large relative to the width of the shock profile. The influence of the beam width is discussed below in describing the reconstruction results for the circular cone. It also acquires errors from interpolation and insufficient accuracy of compensation.

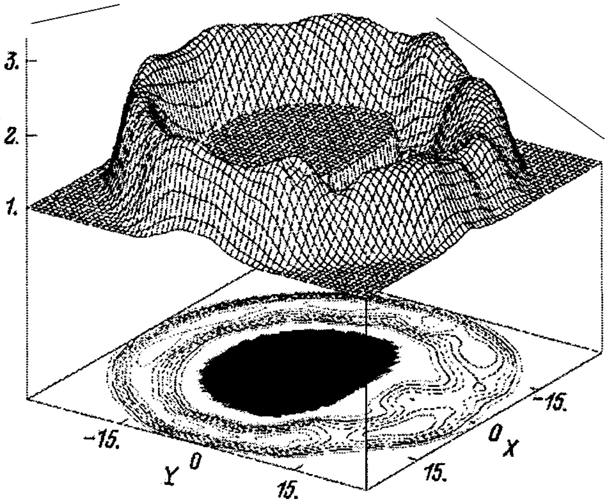

FIG. 5. Reconstruction of the variation of the density field in the plane of the cross section in hypersonic flow around an asymmetrical model (its cross section appears as a "step" in axonometric projection and is represented by the darkened region in isolines (equal-density contours).

For the symmetrical model the reconstruction of the distribution of the variation of the gas density in the flow gives much better results. A single projection with a 0.1-mm step has been recorded in this case. All the other projections necessary for reconstruction by the ART1 algorithm are obtained by multiplication of the existing one. Thus, 18 projections are obtained with 301 readings each. The reconstruction zone is also partitioned into 301×301 pixels. The projections are presmoothed by splines[13] on the assumption that they are distorted by random noise with a variance equal to 3% of the maximum.

The cross section of the reconstructed symmetrical distribution of the density variation at the $x$ axis is shown in Fig. 6 (curve *1*). Curve *3* in this figure has been plotted from the measurement data by electron-beam fluorescence of a rarefied nitrogen flow around a sharp-nosed cone with close values of $M$ and $\mathrm{Re}_1$ (Ref. 15). The measurement data are corrected for the difference in the angles of the model cones. It is evident from Fig. 6 that the reconstructed density profile is shifted toward greater distances from the surface and below profile *3*. The first deviation can be attributed to the influence of blunting of the cone and inaccuracy of the compensating increment (15). The second deviation is associated

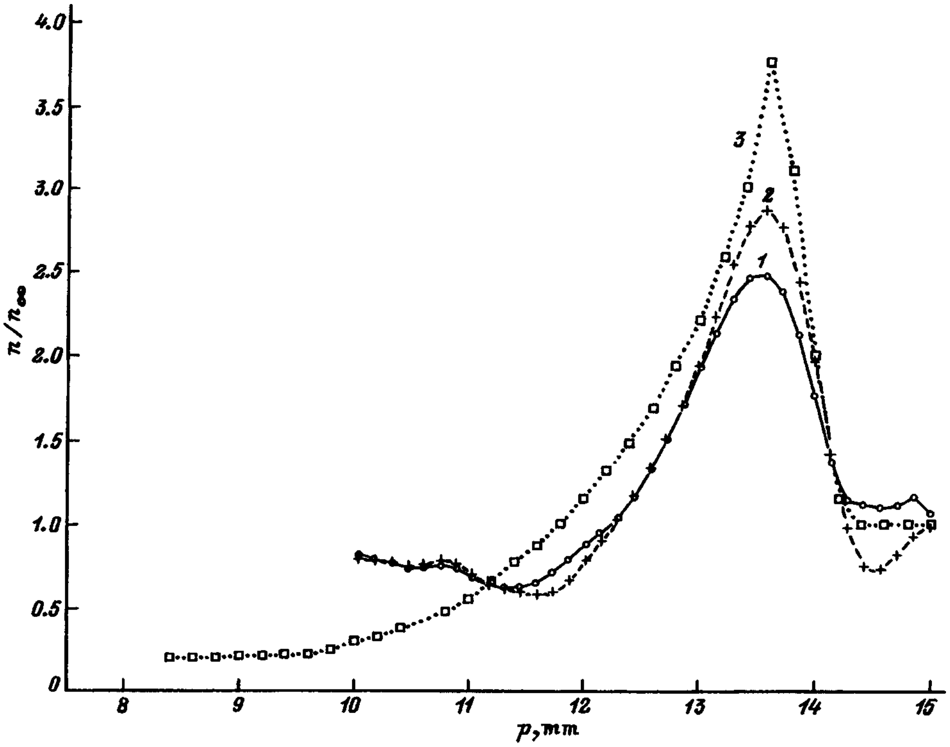

FIG. 6. Cross section, through the $X$ axis, of the reconstructed symmetrical distribution of the density variation in flow around a circular cone.

with greater width of the beam relative to the narrow upper part of the profile. Actually the reconstructed function is averaged over the scales of the beam width, so that a ''sharp'' peak is reconstructed with a lower amplitude. It is important to mention that a reliable procedure for measuring the density of three-dimensional gaseous objects by electron-beam fluorescence under the conditions of Ref. 15 does not exist at the present time, and these data can contain an error that is difficult to estimate.

Obviously, if the tomographic data are collected within the limits of a band having the profile $j(p)$, the resulting parallel projection is the convolution of the projection $f^0(p)$ obtained in infinitely thin rays with the band profile

$$f(p)=\int_{-\infty}^{\infty} f^0(p')j(p-p')dp'. \tag{16}$$

On the other hand, the shape of the outer boundary of an electron beam has been determined experimentally[16] under conditions similar to those of the experiment reported here. It was found that the beam could be accurately regarded as cylindrical in the region of interest to us. In the cross-sectional plane, therefore, the beam represents a band with a known profile. Consequently, the recorded projection is the convolution (16), and the transition to the function $f^0(p)$ corresponds more nearly to the ray tomography approximation.

We have deconvoluted the convolution (16) by a regularizing technique developed in Ref. 17:

$$f^0_*(\nu)=\frac{f_*(\nu)j_*(-\nu)}{|j_*(\nu)|^2+\alpha\nu^2}. \tag{17}$$

The asterisk subscript in Eq. (17) signifies the corresponding Fourier transforms. The regularization parameter $\alpha=0.1$ is selected on the basis of a series of calculations. The result of reconstruction after deconvolution is illustrated by curve *2* in Fig. 6, which shows the cross section of the reconstructed distribution of the density variation at the $x$ axis. It is evident that, as expected, curve *2* agrees better with curve *3* than does curve *1.*

## CONCLUSION

The new measurement procedure and tomographic reconstruction algorithm described in the article have been used to obtain the density distribution in rarefied hypersonic flow around an axisymmetrical object and around an asymmetrical object. In the experimental data preprocessing stage we have solved the problems of interpolating onto a uniform grid and correcting for partial absorption of the electron beam by the model. From data recorded in the region of partial overlap of the beam with the model we have reconstructed for the first time the distribution of the current density in the beam on the assumption of axial symmetry. We have performed deconvolution of the projections with the beam profile determined in the study. We have found that tomographic reconstruction from such data is more accurate. The accuracy of reconstruction can be improved in the future by decreasing the width of the diagnostic beam and increasing the number of sensing aspects for each impact parameter.

The authors are grateful for partial financial support of this study from the Russian Fund for Fundamental Research, Grants No. 95-02-03615 and No. 96-01-01640.

[1] Generally speaking, the discussion in this section is strictly valid as long as scattering does not occur in the flow. However, the influence of this effect is insignificant in the given situation.

[2] In this definition of the elements of the matrix $A$ the reconstructed values of $\Delta n$ are expressed in units of the quantity $\sigma$, but the units are of no

consequence, because the results are normalized to the freestream density.

---

[1] E. P. Muntz, in *Methods of Experimental Physics*, Vol. 18, Part B (Academic Press, New York, 1981), pp. 434–455.

[2] V. N. Vetlutskiĭ, A. A. Maslov, S. G. Mironov *et al.*, Prikl. Mekh. Tekh. Fiz. No. **6**, 60 (1995).

[3] L. I. Kuznetsov and V. N. Yarygin, in *Experimental Methods in Rarefied Gas Dynamics* [in Russian], Inst. Teplofiz. SO AN SSSR, Novosibirsk (1974), pp. 150–165.

[4] F. C. Hurlbut, J. Appl. Phys. **30**, 273 (1959).

[5] É. P. Busygin and G. K. Tumakaev, Zh. Tekh. Fiz. **34**, 122 (1964) [Sov. Phys. Tech. Phys. **9**, 94 (1964)].

[6] I. Wada, in *Rarefied Gas Dynamics*, Vol. 1, edited by J. H. De Leeuw (1965), pp. 535–547.

[7] A. V. Ivanov, Prikl. Mekh. Tekh. Fiz. , No. **6**, 66 (1964).

[8] A. V. Ivanov, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza No. **3**, 108 (1967).

[9] V. V. Pikalov and N. G. Preobrazhenskiĭ, *Reconstructive Tomography in Gas Dynamics and Plasma Physics* [in Russian], Nauka, Novosibirsk, 1987, 231 pp.

[10] I. G. Druker, V. D. Zhak, B. A. Sapogov *et al.*, in *Problems of Gas Dynamics* [in Russian], No. 5 (ITPM SO AN SSSR, Novosibirsk, 1975), p. 31.

[11] G. I. Sukhinin, in *Proceedings of the Sixth All-Union Conference on Rarefied Gas Dynamics* [in Russian] (Novosibirsk, 1979), pp. 79–81.

[12] J. H. Taylor, Astrophys. J. **150**, 421 (1967).

[13] G. I. Marchuk, *Methods of Numerical Mathematics*, 2nd ed. (Springer-Verlag, Berlin-New York, 1986) [Russian orig., Nauka, Moscow, 1977].

[14] G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography* (Academic Press, New York–London, 1980) [Russian trans., Mir, Moscow (1983), 352 pp.].

[15] I. F. Vas and J. G. Sierchio, in *Rarefied Gas Dynamics*, edited by K. Karamchetti (1974), pp. 307–315.

[16] L. H. Clapp, R. J. Catolica, and R. G. Twiss, in *Book of Abstracts of the 17th International Conference on Rarefied Gas Dynamics, Vol. 1* (Aachen, 1990), pp. 319–321.

[17] A. N. Tikhonov and V. Ya. Arsenin, *Solutions of Ill-Posed Problems*, (Halsted Press, New York, 1977; Nauka, Moscow, 1986, 3rd ed. 287 pp.).

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

## BRIEF COMMUNICATIONS

# Temperature of a thermally insulated surface in a gas flow

A. F. Gutsol

*Institute of Chemistry and Technology of Rare Elements and Mineral Raw Materials, Kol'skiĭ Science Center, Russian Academy of Sciences, 184200 Apatity, Murmansk Province, Russia*

(Submitted April 8, 1997)
Zh. Tekh. Fiz. **68**, 134–135 (April 1998)

An equation is derived for determining the temperature of a thermally insulated surface in a gas flow. The equation does not contain any empirical coefficients. The derivation is based on allowance for the work done by a jet arrested at an obstructing surface on the surrounding flow layers. The application of the equation to subsonic and supersonic flows is discussed. © *1998 American Institute of Physics.* [S1063-7842(98)02404-0]

The determination of the temperature of a thermally insulated surface in a high-velocity gas flow is an important problem, one that arises in almost every investigation of the behavior of bodies moving at high velocity relative to the surrounding gas.

The calculation of the surface temperature $T_s$ of a body in a high-velocity gas flow is based on the determination of the flow stagnation temperature $T^*$:

$$T^* = T_0 + 0.5 v_0^2 / c_p . \tag{1}$$

Here $T_0$ is the temperature of the gas, $c_p$ is its specific heat at constant pressure, and $v_0$ is the flow velocity. A correction factor $f$ is sometimes introduced to improve the accuracy of the results,[1] whereupon the surface temperature is written

$$T_s = T_0 + 0.5 f v_0^2 / c_p . \tag{2}$$

The factor $f$ is determined mainly by experimental means (it has an approximate value of 0.8 for subsonic airflow velocities$^{1)}$ or theoretically with allowance for the properties of the boundary layer.[2] This theoretical approach is far from simple, and in the majority of cases it is customary to set $f=1$ in calculations, even in Loĭtsyanskiĭ's fundamental monograph,[2] and also to write

$$T_s = T^* = T_0[1 + 0.5(k-1)M^2]. \tag{3}$$

Here $M$ is the freestream Mach number, and $k = c_p / c_v$ is the specific-heat ratio (a constant) of the adiabatic process. We note that Eq. (3) is simply an alternative representation of Eq. (1), which, in turn, is a special case of the Bernoulli equation. The model of fluid jets with a slowly changing cross-sectional area is used to derive the Bernoulli equation (see, e.g., Ref. 1). However, the impinging jet is abruptly halted when it comes into contact with a surface, inevitably causing it to suddenly spread out. The broadening of the arrested jet induces compression and acceleration of the adjacent unobstructed layers, i.e., the gas of the arrested jet does work, which must be taken into account when the Bernoulli equation is applied to a gas jet arrested in interaction with a surface. By taking into account this work done at a constant pressure equal to the freestream pressure $P$ we can write the Bernoulli equation for one mole of an ideal gas in the form

$$P/\rho_2 - P/\rho_1 + I_2 - I_1 + 0.5(v_2^2 - v_1^2) = 0. \tag{4}$$

Here $\rho$ is the density of the gas, $I = c_p T$ is the enthalpy, and the subscripts 1 and 2 refer to the initial and final states, respectively. Now, making use of the equation of state of the gas

$$P/\rho = R_0 T, \tag{5}$$

along with the full stop of the jet incident on the surface (condition of adherence of the gas to the surface) and the fact that the temperature of the thermally insulated surface is equal to the temperature of the boundary layer formed from the incident gas, we obtain

$$\begin{aligned} T_s &= T_0 + 0.5 v_0^2 / (c_p + R_0) \\ &= T_0[1 + 0.5 M^2 k(k-1)/(2k-1)]. \end{aligned} \tag{6}$$

In this case the correction factor $f = c_p/(c_p + R_0) = k/(2k-1) = 7/9$ for air ($k = 1.4$) is very close to the experimental value.[1]

In the case of a supersonic gas flow the latter is partially braked in the compression shock at the leading edge of the body, and only after that does the gas finally come to rest against the surface of the body. Allowing for the fact that in a normal shock the velocity drops from $v_0$ to $v_1$, the temperature increases from $T_0$ to $T_1$, the density increases from $\rho_0$ to $\rho_1$, and the pressure increases from $P_0$ to $P_1$, where[1]

$$\rho_0/\rho_1 = v_1/v_0 = (k-1)/(k+1) + 2M^{-2}/(k+1), \tag{7}$$

$$P_1/P_0 = 2kM^2/(k+1) - (k-1)/(k+1), \tag{8}$$

we can find the postshock flow temperature:

$$\begin{aligned} T_1 = T_0&[2kM^2/(k+1) - (k-1)/(k+1)] \\ &\times[(k-1)/(k+1) + 2M^{-2}/(k+1)] \end{aligned} \tag{9}$$

and, with the use of Eq. (6), the surface temperature $T_{ss}$ at the leading edge of a body in supersonic flow. The expression for $T_{ss}$ is too cumbersome to write out here. We merely note that the correction factor $f$ in Eq. (2) increases as the Mach number $M>1$ increases, and for air, in particular, it is approximately equal to 0.95 for $M^2=5$ and 0.97 for $M^2=10$. If the body in the flow is sufficiently long, the surface temperature decreases in the sternward direction to $T_s$, because in interaction with the surface the flow acquires segments that have passed through oblique shocks and Mach waves, in other words, that have not undergone as intense prebraking.

The proposed method thus affords a relatively simple means for analytically determining the temperature of a thermally insulated surface in a high-velocity gas flow.

[1] G. N. Abramovich, *Applied Gas Dynamics* [in Russian], Nauka, Moscow (1976), 888 pp.
[2] L. G. Loĭtsyanskiĭ, *Mechanics of Liquids and Gases* [in Russian], Nauka, Moscow (1987), 840 pp.

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Investigation of the parameters of neutron filters

V. G. Kiptilyĭ

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

A. V. Livke, V. I. Nagornyĭ, Yu. Ya. Nefedov, M. V. Savin, V. I. Semenov, and V. A. Chirkin

*Russian Federal Nuclear Center — All-Russian Scientific-Research Institute of Experimental Physics, Sarov, Russia*

(Submitted June 3, 1997)
Zh. Tekh. Fiz. **68**, 136–138 (April 1998)

The quality of suppression of neutrons by lithium hydride $^6$LiH filters for narrow and wide beams from a neutron generator incorporating deuterium and tritium targets is investigated. The experimental data can be used for the design of measurement apparatus operating in high neutron fields. © *1998 American Institute of Physics.* [S1063-7842(98)02504-5]

The issue of how to detect gamma rays against background neutron radiation arises in the solution of a number of problems in experimental physics. As a rule, the gamma-ray detection efficiency of the detectors used in this case is much higher than their neutron detection efficiency, and at the same time filters that effectively suppress neutron radiation are used.

Our goal in the present study is to investigate the quality of suppression of neutrons by lithium hydride $^6$LiH filters and to determine their efficiency in experiments on a thermonuclear reactor of the tokamak type. The investigated neutron filters (F1, F2, and F3) were thin-walled aluminum cylinders of diameter 30 mm and length 300 mm filled with fine-crystalline lithium hydride. The neutron source in the experiment was an NG-150M generator with tritium and deuterium targets, for which the neutron energies were 14.8 MeV and 2.8 MeV, respectively. The absolute neutron yield of the generator was determined by means of an internal monitor, which detected either $\alpha$ particles or protons, depending on the type of target used. To adjust for the effect of migration of the deuteron beam over the target, the neutron intensity was monitored directly at the input to the measurement system by means of a special detector. A diagram of the experimental arrangement is shown in Fig. 1.

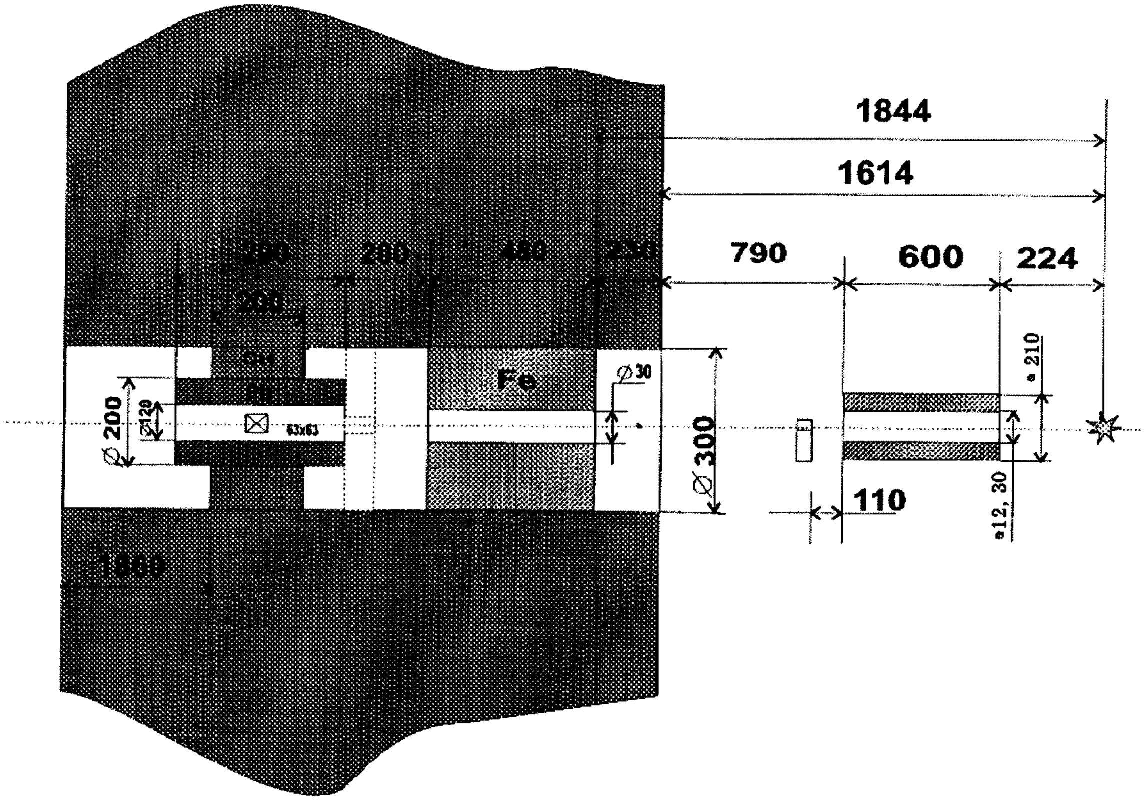

FIG. 1. Geometry of the experiments with $^6$LiH filters on an NG-150M neutron generator.

TABLE I. Attenuation coefficients of 14.8-MeV *DT* neutrons.

| Threshold for | Narrow beam (diam. 12 mm) | | | | Wide beam (diam. 30 mm) | | | |
|---|---|---|---|---|---|---|---|---|
| neutrons, MeV | F1 | F2 | F3 | F1 + F2 | F1 | F2 | F3 | F1 + F2 |
| 10 | 30.6 | 30.6 | 29.6 | 820 | 28.4 | 29.2 | 28.7 | 580 |
| 12 | 29.7 | 28.9 | 29.9 | 810 | 27.0 | 27.8 | 27.4 | 660 |
| 13.5 | 28.6 | 26.2 | 32.5 | 790 | 22.9 | 24.6 | 24.1 | 1200 |

TABLE II. Attenuation coefficients of 2.8-MeV *DD* neutrons.

| Threshold for protons, MeV | F1, wide beam (diam. 30 mm) |
|---|---|
| 2 | 900 |
| 2.4 | 820 |

The neutron detector, surrounded by a lead shield, was placed at a distance of 2.67 m from the target of the generator inside a concrete wall of thickness 3 m. To reduce the neutron flux onto the wall, a protective iron cylinder of diameter 210 mm and thickness 600 mm was placed between the wall and the target of the generator. The first (input) collimator, whose diameter could be set either at 12 mm or at 30 mm, was situated in the middle of the cylinder. The filter irradiation conditions could be simulated by changing the diameter of this collimator, in effect generating either a ''narrow'' or a ''wide'' beam of neutrons. A second iron collimator of thickness 480 mm with a diameter of 30 mm was installed in the concrete wall, and the investigated filters were inserted into it. A monitor was placed between the first and second collimators.

The filter attenuation coefficient is the ratio of the number of neutrons registered by the detector in measurements without a filter to the number of neutrons registered by the detector with the neutron flow intercepted by a neutron filter for the same fluence of neutrons at the entrance to the measurement apparatus. The counting rate of the neutron detector has several components:

$$I_{\mathrm{tot}}=I_n+I_{\gamma G}+I_{\gamma C}+I_{\gamma B},$$

where $I_n$, $I_{\gamma G}$, $I_{\gamma C}$, and $I_{\gamma B}$ are the intensities due to the detection of neutrons, gamma rays from the generator target, gamma rays from slow-neutron capture reactions, and ambient background gamma rays.

The contributions are separated by circuits for the discrimination of gamma-rays and neutrons according to the shape of the light pulse generated in the scintillator. In our work we used a stilbene scintillation neutron spectrometer incorporating a pulse shape discrimination (PSD) system with an adjustable threshold to control the pulse amplitude analysis unit of the spectrometer. The thresholds for the detection of recoil protons were set equal to $\sim 5$ MeV for *DT* neutrons and $\sim 1$ MeV for *DD* neutrons.

The results of neutron measurements are known to depend on the geometry of the experiment, in particular, on the diameter of the neutron beam incident on the sample. In our experiments, therefore, provision was made for testing this effect when working with narrow and wide beams. In the experiments with a narrow (12 mm) beam and a wide (30 mm) beam of 14.8-MeV neutrons we measured the absorption coefficients of single and double (F1 + F2) filters. Only filter F1 was used for the measurements with a wide beam of 2.8-MeV neutrons.

The processing of the measurement results was identical for both cases: For each series of measurements in the PSD control regime we determined the number of pulses recorded by the pulse analyzer above a certain threshold, which corresponded to 10 MeV, 12 MeV, and 13.5 MeV for *DT* neutrons and to 2.0 MeV and 2.4 MeV for *DD* neutrons.

The attenuation coefficient $K$ was determined from the relation

$$K=\frac{N_0}{[N_f-N_B(t_f/t_B)]-[N_{\mathrm{Fe}}-N_B(t_{\mathrm{Fe}}/t_B)](M_f/M_{\mathrm{Fe}})}\,\frac{M_f}{M_0},$$

where $N_0$, $N_f$, and $N_{\mathrm{Fe}}$ are the above-threshold detector counting rates in experiments without a filter, with a $^6$LiH filter, and with an Fe filter, $N_B$ is the background counting rate, $M_0$, $M_f$, and $M_{\mathrm{Fe}}$ are the corresponding monitor readings, and $t_f$, $t_{\mathrm{Fe}}$, and $t_B$ are the exposure times.

Tables I and II give the results of measurements of the attenuation coefficients $K$ in experiments with *DT* neutrons and *DD* neutrons. The error of determination of $K$ in the *DT*-neutron experiments does not exceed 5% for a single filter and 10–15% for two filters at detection thresholds of 10 MeV and 12 MeV. The error in the *DD*-neutron experiments does not exceed 10% for a 2-MeV threshold and 15% for a 2.4-MeV threshold. The values of the attenuation coefficients of filters F1, F2, and F3 agree within the measurements error limits, indicating their uniform filling with the neutron absorber.

Of special interest in tokamak experiments is the estimation of the contribution of capture gamma rays $I_{\gamma C}$ from measurements using an iron filter. We find that in the energy range $>10$ MeV the contribution of gamma rays to the amplitude distribution is 85–30%. One technique for diminishing the contribution of capture gamma rays is to equip the detector with an additional lead shield. To test the effectiveness of this method, we have performed an experiments using two filters (F1 + F2) in a *DT*-neutron wide-beam geometry, in which an additional lead shield of thickness 100 mm was used with the detector, as represented by the dashed line in Fig. 1. The amplitude spectra obtained in this experiment show that the contribution of capture gamma rays in the energy range $>10$ MeV amounts to 30–10%. Consequently, additional shielding of the detector is effective and highly recommended in tokamak measurement apparatus, where the neutron fluxes exceed the gamma-ray fluxes.

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 4 APRIL 1998

# Multidetector device for the detection of coincidences of charged particles and $\gamma$ rays

M. F. Kudoyarov, V. V. Lobanov, and A. A. Pasternak

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

S. I. Lashaev

*V. G. Khlopin Radium Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia*

(Submitted September 10, 1995; resubmitted October 21, 1997)
Zh. Tekh. Fiz. **68**, 139–142 (April 1998)

A $4\pi$ position-sensitive, axisymmetrical assembly of Si–Au charged-particle detectors is proposed, implemented, and tested on a beam of heavy ions; the dimensions and structure of the device are conducive to the organization of coincidences of charged reaction products with discrete $\gamma$ rays emitted by the daughter nucleus and registered by a system of ultrapure Ge detectors. First results are obtained from an investigation of the reaction $^{58}$Ni($^{16}$O,$\alpha 2p\gamma$)$^{68}$Ge at $E_0 = 74.5$ MeV. © *1998 American Institute of Physics.* [S1063-7842(98)02604-X]

## INTRODUCTION

The acquisition of new information in low-energy nuclear physics largely entails the development of physical apparatus, especially detection systems for such apparatus. In the last decade, thanks to the efforts of several international groups, a number of large-scale devices have been built for the detection of products of nuclear reactions involving low-energy and medium-energy heavy ions, specifically $\gamma$ rays and charged particles. Present-day multicrystal $\gamma$-ray spectrometers of the ''crystal ball'' type, consisting of $4\pi$ assemblies of decades of ultrapure Ge single crystals partially augmented with $4\pi$ BGO $\gamma$-ray multiplicity filters, are designed primarily for studies of the physics of high-spin states of atomic nuclei.[1–6] The mechanisms of nuclear reactions with heavy ions are investigated mainly by multidetector position-sensitive devices consisting of decades and even hundreds of charged-particle detectors employing diverse operating principles. Some are devices of the FOBOS type,[7] designed for the detection of light charged reaction products and fission fragments, or mass separators of recoil nuclei. However, the large overall dimensions of these devices (measuring in meters) make it impossible to organize the coincidences of $\gamma$ rays with particles and, hence, to investigate a major category of problems that have important bearing both on the physics of high-spin states and on our understanding of the mechanisms underlying nuclear reactions.

The only real potential lies in the design of small-scale ($\leqslant 0.1$ m) charged-particle detection systems that can be housed in $\gamma$-ray spectrometers. In principle, it should be possible to build a semiconductor detector in the form of a segment of a sphere and to configure a $4\pi$ geometry from such parts, but the requirement of coordinate resolution, no matter how coarse, together with the specific form of the detector itself poses an exceedingly complex technological problem. Well-known attempts have been made to construct assemblies of planar semiconductor detectors in the shape of a hollow cube (the OSIRIS Cube[9]) and to use them in $\gamma$-ray spectrometers for the identification of reaction channels in which protons or $\alpha$ particles are emitted.[10] The drawbacks of such a system include: 1) sizable dead zones at the detector junctions; 2) a disparity of conditions for the detection of particles emanating from a target at an identical angle relative to the beam axis but impinging on the detector plane at different angles, thereby degrading energy resolution; 3) limited coordinate-resolution capabilities, particularly in the range of small forward angles; any such capabilities are achievable only by stacking a number of planar elements (DIAMAND) and, accordingly, increasing the number of electron channels.

## STRUCTURE OF THE DEVICE

We have proposed and implemented an axisymmetrical, multidetector device of a different type, which provides not only a $4\pi$ charged-particle detection geometry, but also the possibility of polar-angle coordinate resolution. The advantages of such an assembly, which consists of cylindrical and disk elements, are: 1) a simpler element fabrication technology (in contrast with spherical geometry); 2) identical particle detection conditions with respect to azimuth angle, a feature not found in hollow-cube assemblies; 3) the possibility of producing several independent detecting rings on a single crystal, providing the means for pinpointing the particle exit angle from a target within a narrow angular interval; 4) reduction of the number of electron spectrometer channels for cylindrical elements by one fourth compared with the number required to cover the corresponding geometry in a system of the ''box'' type; 5) the possibility of reducing the width of the dead zones between adjacent detectors to 1 mm or less.

We have developed and run certification tests on a technology for the fabrication of Si–Au detectors, each in the form of a thin-walled ($\approx 1$ mm) cylinder with the sensitive layer lining its inner surface. We have also worked out the technology for constructing several independent detectors on one silicon crystal without any mutual influence between detectors. The dead zones between them do not exceed 1 mm.

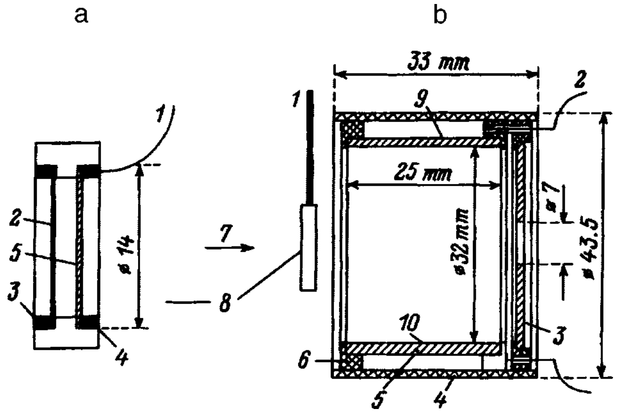

FIG. 1. Fragment (forward part) of the reaction chamber in cross section. a) Blowup of the target device: *1*) contact lead; *2*) target; *3*) dielectric ring; *4*) metal ring; *5*) Ta foil ($3\times8=24$ $\mu$m). b) chamber proper: *1*) target holder; *2*) contact lead; *3*) Si ring (sensitive layer $\approx500$ $\mu$m); *4*) Al jacket; *5*) Si cylinder (sensitive layer $\approx300$ $\mu$m); *6*) insulator; *7*) beam; *8*) target unit; *9*) Ni layer; *10*) Au layer.

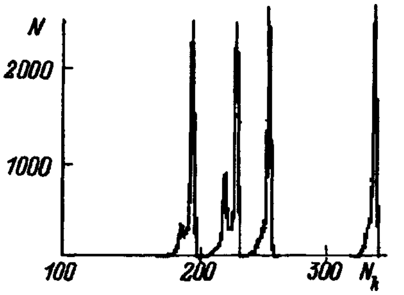

FIG. 2. Spectrum of $^{226}$Ra $\alpha$ particles, obtained from a cylindrical Si–Au detector.

The cylinder is made from a high-resistance *n*-Si single crystal by coring out its interior. The subsequent processing of the surface and the formation of a working zone are similar in technology to standard methods used in the fabrication of conventional Si–Au detectors. Figure 1 shows a fragment of the device in cross section, including the semiconductor detectors.

The beam enters through an aperture in the rear annular detector of diameter 8 mm. The target is placed on an autonomous holder between two identical cylinders in such a way as to minimize the gap between the ends of the cylinders. The target holder is designed so that a heavy metal foil (tantalum in our case) if placed directly behind the target (in the direction of the beam) to shield the detector against the direct incidence of beam particles that pass through the target without interaction. The cylinder ends far from the target are capped with interchangeable annular Si–Au detectors with a thick (up to 500 $\mu$m) working zone, which ensures the total absorption of $\alpha$ particles with energies $\approx$40 MeV. Various types of annular detectors, in particular a detector consisting of three axisymmetrical rings formed on one crystal, are used in our experiments. This arrangement makes it possible to segregate reaction products emanating from the target in well-defined intervals of the exit angle and, combined with energy measurement, to determine the reaction products with the required accuracy and, hence to distinguish the channels necessary for further processing. The area of the working surface of the cylindrical detector is 20 cm$^2$, and its capacitance is $\approx$1000 pF, requiring the use of special preamplifiers.[11]

As an illustration, Fig. 2 shows the spectrum of $^{226}$Ra $\alpha$ particles obtained using a cylindrical detector. The source is placed at the target site, i.e., the incidence of $\alpha$ particles on the detector is "oblique," further degrading the energy resolution by stretching out the left edge of the peak. All the same, $\approx$70-keV resolution is obtained in the given situation, which is fully acceptable for practical applications.

In our first experiments, primarily of a methodological and demonstrative orientation, two Canberra $\gamma$-ray detectors already at our disposal, made from germanium of extreme purity with a volume of 90 cm$^3$ each, were used instead of a "crystal ball" for $\gamma$-ray detection. A block diagram of the device is shown in Fig. 3. The detector output signals branch out after preamplification: The fast signal (with a 10-ns leading edge) is sent from a special *T*-output to a fast standard-pulse shaper and then to a coincidence circuit; the spectrometric signal, amplified and shaped by a linear amplifier, is sent to an analog-to-digital converter (ADC). The multiple-input coincidence circuit can be used to organize coincidences of the signals in essentially any combination. The coincidence circuit signal, broadened to 5 $\mu$m by a univibrator, opens the ADC gates.

This arrangement ensures that spectrometric information is recorded only when the detectors are triggered simultaneously in a combination predetermined by the coincidence circuit, making it possible to effectively identify the reaction channels of interest and, by adjustment of the lower and upper thresholds of the amplifiers, to separate out the necessary energy intervals.

The electron channels of the Ge and Si–Au detectors are unified, the only difference being in the type of converters used. For the Ge detectors, 4096-channel converters with enhanced linearity and stability are used, so that an energy resolution of 2.0–2.1 keV can be maintained for the duration of the experiment (several days) by using Canberra spectrometric amplifiers.

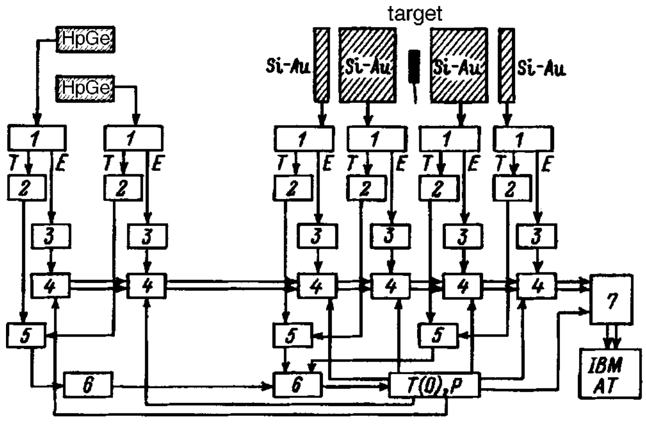

FIG. 3. Block diagram of coincidence detection. *1*) Preamplifier; *2*) shaping circuit; *3*) linear amplifier; *4*) ADC; *5*) logic-pulse mixer; *6*) coincidence circuit; $T(0),P$) univibrator; *7*) CAMAC crate controller.

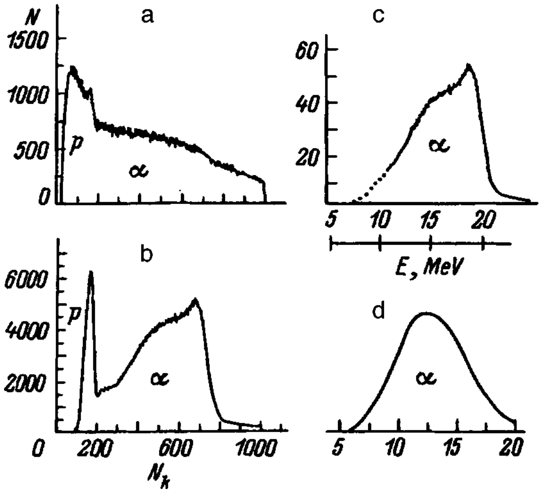

FIG. 4. a) Direct spectrum of charged particles from the forward Si–Au detector; b) spectrum of ternary $\gamma$-Si–Au forward–Si–Au cylinder coincidences; c) spectrum in the center-of-mass system; d) calculated spectrum of evaporative $\alpha$ particles.

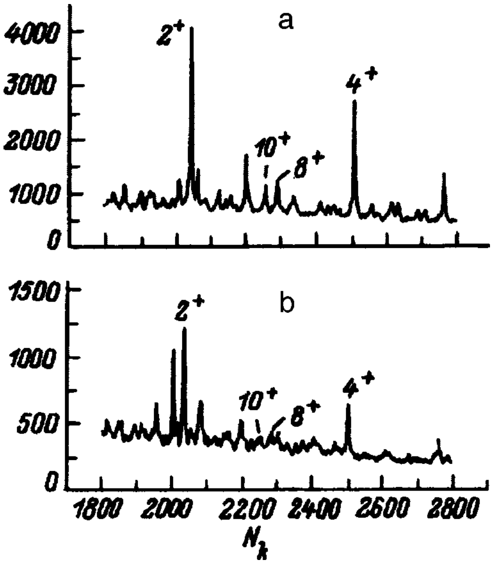

FIG. 5. a) $\gamma$-Ray spectrum in coincidence with the forward Si–Au detector; b) direct $\gamma$-ray spectrum.

## FIRST EXPERIMENTAL RESULTS

In our first experiments using a heavy-ion beam from the cyclotron at the A. F. Ioffe Physicotechnical Institute of the Russian Academy of Sciences (FTI) we investigated the spectra of coincidences of $\gamma$ rays with $\alpha$ particles and protons in the reaction $^{58}\mathrm{Ni}(^{16}\mathrm{O},\alpha 2p\gamma)^{68}\mathrm{Ge}$ at $E_0 = 74.5$ MeV (the choice of this reaction was dictated by the fact that in the FTI cyclotron laboratory the technique of coincidences of $\gamma$ rays with $\alpha$ particles registered by a conventional Si–Au detector had already disclosed the nonequilibrium character of the $\alpha$-particle emission,[12] and these experiments were the stimulus for development of the device described in the present article). The bombarding particles were stopped by a Ta foil of thickness $\approx 40$ mg/cm$^2$, which permitted the passage through it of $\alpha$ and $p$ emitted in the forward angles. Under our experimental conditions $\alpha$ and $p$ are separated by virtue of the difference in the detected energy spectra and angular distributions. Consequently, by sorting out events associated with the incidence of $\alpha$ particles in the forward disk detector and $p$ in the two cylindrical detectors, it is possible to identify the channel of the reaction $(\alpha, 2p)$ and, by coincidences with $\gamma$ rays of the final nucleus, to estimate the fraction of the cross section and the profile of the particle spectrum corresponding to the formation of the double nuclear system. A total of $\approx 2.5 \times 10^6$ events were stored in the experiment.

Omitting the physical interpretation of the results within the space limitations of the article, we can still draw two basic conclusions.

1. It is evident from Fig. 4 that the spectra of $\alpha$ particles detected in coincidence with $\gamma$ rays differ sharply both from the direct spectrum containing the high-energy contribution of $^{16}$O decay and from the quasievaporative spectrum (in the form of a Maxwell distribution with maximum in the vicinity of 12–15 MeV), as is typical of the incomplete fusion mechanism.

2. As illustrated qualitatively in Fig. 5, there is a significant change (relative to the direct spectrum) in the nature of the population of high-spin states of $^{68}$Ge in the selection of events corresponding to coincidence with nonequilibrium $\alpha$ particles emitted in the forward angles as a result of the incomplete fusion mechanism.

## PROSPECTS FOR APPLICATIONS OF THE SYSTEM IN THE CONFIGURATION OF MULTIDETECTOR $\gamma$-RAY SPECTROMETERS

It is evident from the brief description of the structure of the device and the results of first experiments that we are now in a position to plan a set of problems for applications in devices of the ball type. The investigation of the mechanism of incomplete fusion of nuclei accompanied by the forward ejection of a fast $\alpha$ particle and the formation of a rapidly spinning, quasinonequilibrium residual nucleus emitting light particles, and then $\gamma$ rays, is a timely subject for future research from the standpoint of the physics of nuclear reactions. The selection of events associated with discrete $\gamma$-ray transitions of the final nucleus in a narrow range of spins affords the possibility of uniquely determining not only the final reaction channel, but also the transferred angular momentum associated with the impact parameter of peripheral collisions. This capability will help to shed light on many obscure problems in the physical of mass transfers in the interaction of heavy ions. On the other hand, individual bands of high-spin states of the daughter nuclei can be identified by sorting out $\gamma$ rays corresponding to charged-particle detection events associated with the incomplete fusion mechanism in peripheral collisions, thereby gaining added possibilities for investigating them by $\gamma$-ray spectroscopy methods. In particular, an additional tool is found for seeking out new regions of superdeformation and hyperdeformation in light nuclei, for which the cross sections of formation in reactions with heavy ions accompanied by charged-particle emission are fairly large in comparison with the neutron channels.

The range of prospective problems for the newly designed device is not exhausted by the study of the coincidence spectra of charged particles with discrete $\gamma$ rays. In particular, in our first collaborative experiment with colleges of the Joint Institute for Nuclear Research we have recorded the spectrum of high-energy $\gamma$ rays detected by a BGO crystal in coincidence with $\gamma$ particles in the reaction $^{58}$Ni($^{16}$O,$\alpha 2p\gamma$)$^{68}$Ge at $E_0 = 74.5$ MeV (Ref. 13). The first results indicate an encouraging outlook for the application of our device in studies of giant resonances using BGO assemblies (of the ball type in particular).

The authors are grateful to the Russian Fund for Fundamental Research for financial support of our work.

[1] *Gammasphere: A National Gamma-Ray Facility, PUB-5202*, edited by M. A. Deleplanque and R. M. Diamond (Berkeley, 1988).

[2] R. M. Diamond, in *Instrumentation for Heavy-Ion Nuclear Research*, edited by D. Shapira (New York, 1985), p. 259.

[3] J. F. Sharpey-Schafer and J. Simpson, Part. Nucl. Phys. **21**, 293 (1988).

[4] B. Herskind, Nucl. Phys. A **447**, 395 (1985).

[5] *Eurogam* (CRN document) (Centre de Recherches Nucleaires, Strasbourg, 1990), pp. 1–72.

[6] J. P. Martin *et al.*, Nucl. Instrum. Methods Phys. Res. A **257**, 301 (1987).

[7] H.-G. Ortlep *et al.*, in *Heavy Ion Physics*, Vol. 2, edited by Yu. Ts. Oganessian (Dubna, 1993), pp. 466–477.

[8] A. R. Sinha *et al.*, Nucl. Instrum. Methods Phys. Res. A **339**, 543 (1994).

[9] R. Schwengner *et al.*, in *Annual Report 1991, FZR 92-09* (Forschungszentrum Rossendorf, Dresden, 1991), p. 37.

[10] G. Winter *et al.*, Z. Phys. A **344**, 229 (1992).

[11] V. N. Dukhanov, N. V. Zhernov, and I. B. Mazurov, Vopr. At. Nauki Tekh. Ser. Obshch. Yad. Fiz., No. 12(42), 87 (1988).

[12] A. E. Antropov, M. F. Kudoyarov *et al.*, in *Proceedings of the 38th Conference on Nuclear Spectroscopy and the Structure of the Atomic Nucleus* [in Russian] (Nauka, Leningrad, 1988), p. 384.

[13] M. F. Kudoyarov, V. V. Lobanov *et al.*, in *Proceedings of the International Conference on Nuclear Spectroscopy and the Structure of the Atomic Nucleus* [in Russian] (St. Petersburg, 1995), p. 384.

Translated by James S. Wood