Modeling and calculation of the capacitance of a planar capacitor containing a ferroelectric thin film

O. G. Vendik, S. P. Zubko, and M. A. Nikol'skiĭ

St. Petersburg State Electrical Engineering University, 197376 St. Petersburg, Russia
(Submitted December 24, 1997)
Zh. Tekh. Fiz. **69**, 1–7 (April 1999)

The capacitance of a two-layer planar capacitor containing a thin layer of $SrTiO_3$ is calculated by conformal mapping using the partial capacitanc e method. Simple formulas are obtained for approximation calculation of the capacitances of individual components of a planar structure, and their limits of applicability are determined. A relation for the capacitance of a planar capacitor is derived which takes account of the size effect in a ferroelectric film within the context of the partial capacitance method. The calculated result is compared with the experimentally measured capacitance. © *1999 American Institute of Physics.* [S1063-7842(99)00104-X]

INTRODUCTION

Recently interest has been rekindled in the practical use of controllable ferroelectric components in microwave technology.[1–4] This has been largely the result of advances in microwave cryoelectronics at liquid-nitrogen temperatures stimulated by developments in microwave circuit elements based on high-temperature superconductors.[5] The practical application of devices working at liquid-nitrogen temperatures makes it possible to actively employ the ferroelectric nonlinearity, since the dielectric losses at this temperature are much lower than at room temperature.[6] The large value of the dielectric constant of ferroelectrics has led to the necessity of using these materials in the form of thin films with thickness on the order of one to several microns when used as components of microwave devices.[6] Thus, the use of ferroelectrics in microwave devices is based on a planar technology in which the main design elements are a planar and a capacitor, a coplanar or gap transmission line. To utilize these elements toward this end, it is necessary to develop mathematical models of these elements which can form the basis of a system of computer-aided design (CAD) of microwave devices utilizing ferroelectric materials. The development of such models is the goal of the present work.

The dependence of the dielectric constant of bulk and film samples of ferroelectric materials on the temperature and electric field strength is well described on the basis of a validated phenomenological model.[7,8] In addition, models are needed which allow one to calculate the capacitance of a planar structure formed by a ferroelectric thin film on an insulating substrate with metal (superconducting) planar electrodes with allowance for the fringing fields in the ambient free space (air). Such a calculation can be based on the method of conformal mapping, which allows one to transform the field in a planar structure into the field of a sandwich capacitor.[9] The planar structure in question consists of three component parts: the ferroelectric film, substrate, and air. A separate calculation of the capacitance of each of the component parts with subsequent summation constitutes the basis of the partial capacitance method. This method was first implemented in the works of Kochanov;[10,11] however, its current name was gradually formed only in the English-language literature.[12–16] The partial capacitance method is based on the introduction of zero boundary conditions

for the normal components of the field at the boundary of the media — the ''magnetic wall'' — and subsequent correction of the value of the dielectric constant of each of the media. The use of this method to calculate complicated planar structures requires an estimate of its accuracy. Such an estimate was first made for a coplanar transmission line[11] by comparing with a calculation performed on an analog computer: the error did not exceed 1–2%. Later the method was included in a handbook[12] and used in calculations as *a priori* reliable without an estimate of possible errors.[13–15] A comparison was made in Ref. 16 of the calculated and experimentally measured group velocity of the wave in a coplanar line. The revealed error did not exceed 5%, which can most probably be assigned to errors in the measurements. For a two-layer planar structure containing a thin film with a large dielectric constant ($\varepsilon = 300–10\,000$), and a thick substrate ($\varepsilon = 10–25$), the applicability of the partial capacitance method was subjected to a special analysis,[17] which made it possible to determine the limits of applicability of the method. Corresponding quantitative estimates are given below.

CONFORMAL MAPPING OF A PLANAR STRUCTURE

The distribution of the electric field in planar capacitors is nonuniform.

To calculate the capacitance of a planar capacitor, it is customary to use conformal mapping. The use of conformal mapping based on the Christoffel–Schwarz transformation[9] makes it possible to transform the structures shown in Fig. 1a and 1b into a rectangle (Fig. 1c). The capacitance of a sandwich capacitor without fringing fields (Fig. 1c) is easily calculated:

$$C = \frac{\varepsilon_0 \varepsilon_r K(k')}{2K(k)}, \quad k' = \sqrt{1-k^2}, \tag{1}$$

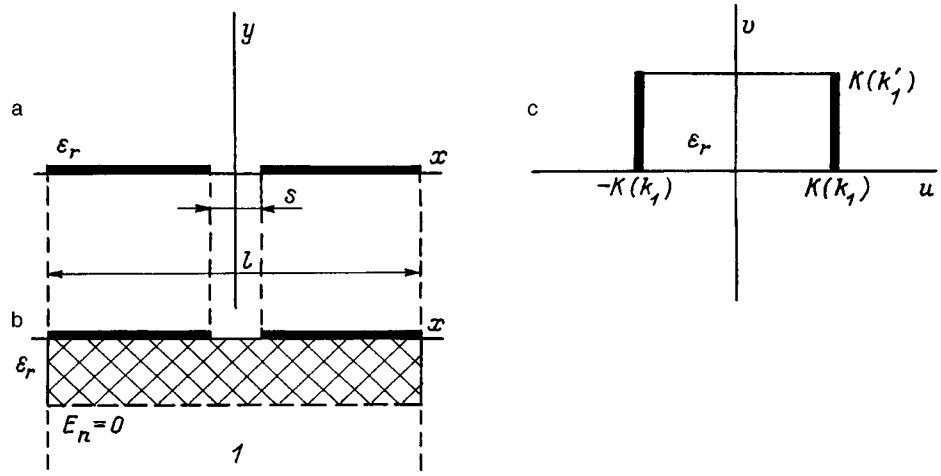

FIG. 1. Conformal mapping of the planar structure: a — two planar electrodes in free space; b — result of mapping into a sandwich capacitor without fringing fields.

where $K(k)$ is the total elliptic integral of the first kind, k is the modulus of the elliptic integral, ε_0 is the permittivity of free space, and ε_r is the dielectric constant of the medium in which the electric field is concentrated.

For the structures shown in Fig. 1a and 1b we have

$$k_a=\frac{s}{l}, \quad k_b=\tanh\left(\frac{\pi s}{4h}\right)/\tanh\left(\frac{\pi l}{4h}\right). \tag{2),(3}$$

The ratio $K(k')/K(k)$ can be represented by the following approximate form (Ref. 18, pp. 38 and 39):

$$F(k)=\frac{K(k')}{K(k)}=\begin{cases}\pi^{-1}\ln\left[2\dfrac{1+(1-k^2)^{0.25}}{1-(1-k^2)^{0.25}}\right] & \text{for } k^2\leqslant 0.5,\\ \pi\left[\ln\left(2\dfrac{1+k^{0.5}}{1-k^{0.5}}\right)\right]^{-1} & \text{for } k^2\geqslant 0.5.\end{cases} \tag{4}$$

The error of the above approximate form does not exceed 10^{-5} (Ref. 18). Note that the above formulas give the capacitance of planar structures per unit length in the direction perpendicular to the plane of the figure.

FORMULAS FOR CALCULATING THE CAPACITANCE OF THE COMPONENT PARTS OF A PLANAR STRUCTURE

Consider a planar capacitor (Fig. 2). Its capacitance is formed by the capacitance of the stray fields in the ambient space (air) with dielectric constant $\varepsilon_1=1$, the capacitance of the ferroelectric layer of thickness h_2 with dielectric constant ε_2, and the capacitance of the substrate of thickness h_3 with dielectric constant ε_3. The calculation was based on the partial capacitance method.[10–16] We represent the composite layered capacitor (Fig. 2) as three simple planar capacitors with homogeneous filler (Fig. 3) connected in parallel, and seek its capacitance as the sum of the three partial capacitances

$$C=C_1+C_2+C_3, \tag{5}$$

where C_1, C_2, and C_3 are the capacitances of the component parts of the planar capacitor, which are the fringing field in air, the ferroelectric film, and the substrate.

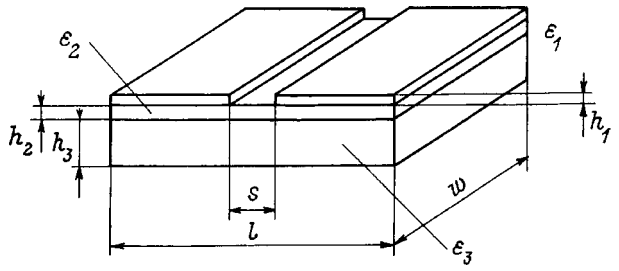

FIG. 2. Diagram of a planar capacitor consisting of an insulating substrate, a thin dielectric layer with a large dielectric constant and conducting electrodes separated by a gap of width s.

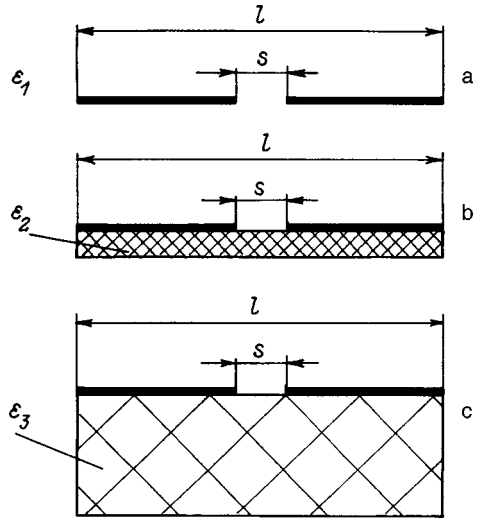

FIG. 3. Separation of a planar capacitor (Fig. 2) into partial capacitances (the procedure underlying the present calculation): a — electrodes in ''air''; b — thin dielectric layer, c — thick dielectric layer.

Here the dielectric constants of the media are decreased in accordance with the equalities

$$\varepsilon_2^* = \varepsilon_2 - \varepsilon_3, \quad \varepsilon_3^* = \varepsilon_3 - 1. \tag{6}$$

These changes in the values of the dielectric constants of the layers constitute the main idea of the partial capacitance method, in which each layer is screened from the other layers by a ''magnetic wall'' which does not transmit the normal components of the electric field vectors. The mutual influence of the various layers on the field distribution in each of them is taken into account by the changes in the values of the dielectric constant of each of the layers represented in the case under consideration by relations (6).

Now let us calculate each capacitance separately taking the above into account.

1. The capacitance of the air gap (Fig. 3a)

$$k_1 = \frac{s}{l}, \quad C_1 = \varepsilon_0 F(k_1) \cdot w. \tag{7}$$

2. The capacitance of the ferroelectric layer bounded by the ''magnetic wall'' (Fig. 3b)

$$k_2 = \tanh\left(\frac{\pi s}{4h_2}\right) / \tanh\left(\frac{\pi l}{4h_2}\right), \quad C_2 = 0.5\varepsilon_0 \varepsilon_2^* F(k_2) \cdot w. \tag{8}$$

3. The capacitance of the substrate bounded from below by the ''magnetic wall'' (Fig. 3c)

$$k_3 = \tanh\left(\frac{\pi s}{4(h_3+h_2)}\right) / \tanh\left(\frac{\pi l}{4(h_3+h_2)}\right),$$

$$C_3 = 0.5\varepsilon_0 \varepsilon_3^* F(k_3) \cdot w. \tag{9}$$

Formulas (7)–(9) allow for the width of the planar capacitor w indicated in Fig. 2. For practical calculations it is convenient to use a simplified variant of formulas (7)–(9), which is obtained by expanding the given formulas in a series in a small parameter. As a result we obtain

$$C_1 = w \cdot \varepsilon_0 \frac{2}{\pi} \ln\left(4\frac{l}{s}\right), \tag{10}$$

$$C_2 = \frac{w \cdot \varepsilon_0 (\varepsilon_2 - \varepsilon_3)}{s/h_2 + (4/\pi)\ln 2}, \tag{10'}$$

$$C_2 = w \cdot \varepsilon_0 (\varepsilon_3 - 1) \frac{1}{\pi} \ln\left(16\frac{h_3 - h_2}{\pi s}\right). \tag{10''}$$

Formulas (10) use the quantities sll, h_2/s, and s/h_3, respectively, as the small parameter.

LIMITS OF APPLICABILITY OF THE CALCULATIONAL FORMULAS

The partial capacitance method has a bounded region of applicability. The reason is that the partial capacitance method is based on introducing

''magnetic walls'' between separate layers of the planar structure. Some correction is achieved by the correction of the dielectric constant of the layers. For a large distance between the electrodes the electric field lines are no longer concentrated within the chosen layer but extend beyond it, thereby violating the concept of the ''magnetic wall.'' It is possible to formulate the following restriction on the applicability of the method, which vouchsafes the applicability of the concept of a ''magnetic wall.''

The method is applicable if the inequality

$$s \leqslant 10h_2 \tag{11}$$

is satisfied for $\varepsilon_2/\varepsilon_3 > 10^2$.

This is in good agreement with results[17] calculated for a two-layer planar structure without assuming the presence of a magnetic wall. Calculation shows that if inequality (11) is satisfied the error in the calculation of the total capacitance of the planar structure arising as a consequence of the above approximations will not exceed $\pm 1\%$.

We may also compare the results obtained using the exact formulas obtained by conformal mapping, (7)–(9), and using the simplified formulas (10)–(10″).

We consider a specific realization of a planar capacitor in which we vary the width of the gap s over wide limits for the remaining dimensions fixed (these are given in Table I).

Figures 4a–4c plot the values of each of the partial capacitances calculated according to formulas (5)–(7) and (8)–(10). For ease of visualization, the inverse capacitance of each of the layers is plotted as a function of the relative width of the gap. In each of these cases the gap width is scaled to a characteristic length of the corresponding structure: in the case of electrodes it is their length, in the case of dielectric layers it is their thickness. It can be seen from the figure that the simplified formulas work well if the following inequalities are satisfied: 1) for the capacitance formed by the fringing fields in air

$$s \leqslant 0.25l; \tag{12}$$

2) for the capacitance of the ferroelectric layer

$$s \geqslant h_2; \tag{13}$$

3) for the capacitance of the substrate

$$s \leqslant 0.5h_3. \tag{14}$$

Obviously, inequalities (12)–(14) are to be applied in conjunction with inequality (11).

ACCOUNT OF THE SIZE EFFECT IN THE FERROELECTRIC FILM

It is well known[8] that a size effect is manifested in thin films of ferroelectric materials, i.e., they manifest the dependence of the electrophysical parameters of the material on the thickness of the layer.

It has been established that the size effect manifests itself differently

TABLE I. Numerical characteristics of planar structures for calculation of the partial capacitances, results of which are plotted in Figs. 4a–4c.

	Thickness, μm	ε_r	Length l, μm	Width w, μm
Electrodes	$h_1 = 0$	⋯	2000	$w = 600$
Layer 2	$h_2 = 1$	$\varepsilon_2 = 1000$	2000	$w = 600$
Layer 3	$h_3 = 500$	$\varepsilon_3 = 11$	2000	$w = 600$

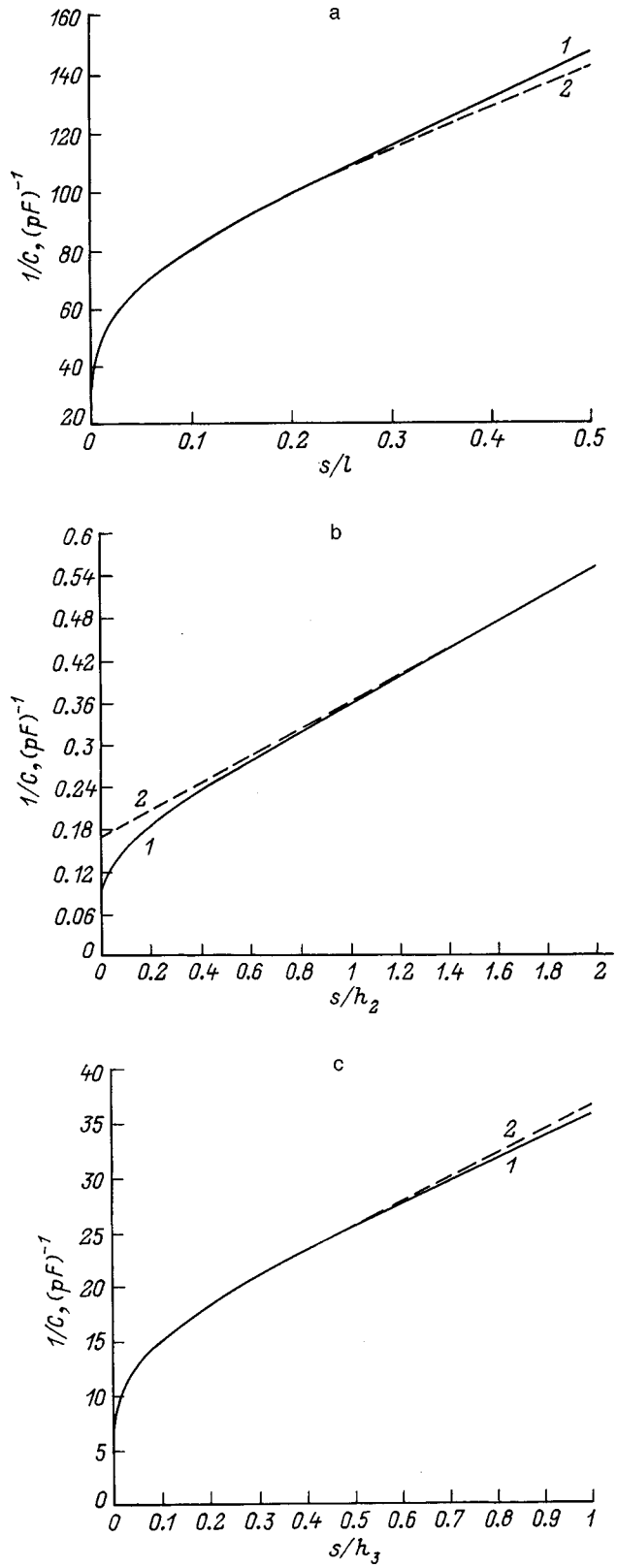

FIG. 4. Limits of agreement of the calculation of partial capacitances based on the exact formulas and on simplified formulas obtained by conformal mapping: a–c — the partial capacitances shown in Fig. 3.

for different orientation of the ferroelectric polarization vector relative to the boundaries of the layer.[19] Here the form of the boundary conditions plays a role: metal, high-temperature superconductor, insulator.[20] We will assume that

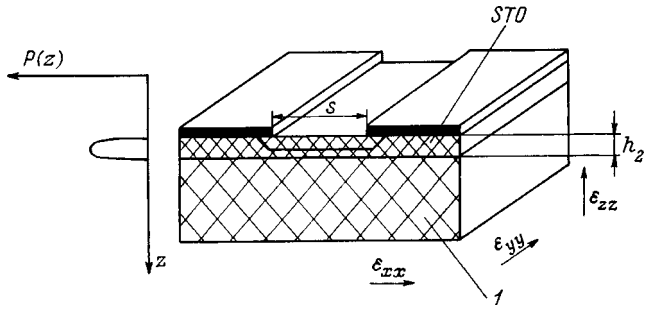

FIG. 5. Formation of effective anisotropy of the ferroelectric layer due to the size effect (*1* is the substrate).

the ferroelectric polarization is equal to zero (''frozen'') on both surfaces of the ferroelectric layer.

Thanks to this, there exists a ferroelectric polarization gradient transverse to the layer. Thus, in the case of a planar capacitor the direction of the electric field vector in the central part of the layer (where the field is uniform) is perpendicular to the polarization gradient (Fig. 7). In this case the size effect can be manifested only if $h<10^{-2}\ \mu$m (Ref. 19). Therefore, with reference to the field components parallel to the boundaries of the layer, the dielectric constant of the ferroelectric is not affected by a size effect.

We denote the dielectric constant relative to the field components parallel to the boundaries of the layer by ε_s. Since the size effect is not manifested relative to these components,

$$\varepsilon_s=\varepsilon_2. \tag{15}$$

We denote by ε_h the effective dielectric constant for the field components directed along the z axis (Fig. 5). Assuming that the ferroelectric polarization is ''frozen'' at the boundaries between layers, we obtain

$$\frac{1}{\varepsilon_h}=\frac{1}{\varepsilon_2}+\frac{2}{\alpha h_2}, \tag{16}$$

where $\alpha=1/\sqrt{2\lambda_1}$ and $\lambda_1=10^{-19}$ (m^2).[8,19,20]

Thus, the dielectric constant of the ferroelectric layer is anisotropic.

The effective anisotropy of the ferroelectric layer due to the size effect can be represented as a tensor

$$|\varepsilon|=\begin{vmatrix}\varepsilon_{xx} & 0 & 0\\ 0 & \varepsilon_{yy} & 0\\ 0 & 0 & \varepsilon_{zz}\end{vmatrix}, \tag{17}$$

where $\varepsilon_{xx}=\varepsilon_{yy}=\varepsilon_2$ and $\varepsilon_{zz}=\varepsilon_h$.

We now apply the formula for the capacitance of a planar capacitor on an anisotropic dielectric layer[9] and obtain

$$C_2=\varepsilon_0\frac{w}{\sqrt{\varepsilon_{yy}}}\left[\frac{s/\sqrt{\varepsilon_{xx}}}{h_2/\sqrt{\varepsilon_{zz}}}+\frac{4}{\pi}\ln 2\right]^{-1}\sqrt{\varepsilon_{xx}\varepsilon_{yy}\varepsilon_{zz}}. \tag{18}$$

Substituting the values of the components of the dielectric tensor of the layer, we obtain an expression for the capacitance of a planar ferroelectri c layer with allowance of the size effect

TABLE II. Mean values of the model parameters obtained for $SrTiO_3$ samples.

Model parameter	ε_{00}	T_C, K	T_F, K	E_N, kV/cm	ξ_s	$C=\varepsilon_{00}T_C$, K
Mean value for a single crystal	2100	42	175	19.0	0.02	8.8×10^4
Mean value for films	3000	34.5	155	7.0	1.2	10.4×10^4

$$C_2=\varepsilon_0\frac{w\varepsilon_2}{s/h_2+(4/\pi)\ln 2-\sqrt{\varepsilon_2/\varepsilon_h}}, \tag{19}$$

where the ratio $\varepsilon_2/\varepsilon_h$ taking relation (16) into account has the form

$$\frac{\varepsilon_2}{\varepsilon_h}=1+\frac{2\varepsilon_2}{\alpha h_2}. \tag{20}$$

Thus, the capacitance with the size effect taken into account within the framework of the partial capacitance method can finally be represented in the form

$$C=\varepsilon_0 w\cdot\left[\frac{2}{\pi}\ln\left(\frac{4l}{s}\right)+\frac{\varepsilon_3-1}{\pi}\ln\left(16\frac{(h_3+h_2)}{\pi s}\right)+\frac{\varepsilon_2-\varepsilon_3}{s/h_2+(4/\pi)\ln 2\sqrt{1+2\varepsilon_2/\alpha h_2}}\right]. \tag{21}$$

MODELING OF THE DIELECTRIC CONSTANT OF A FERROELECTRIC LAYER AS A FUNCTION OF TEMPERATURE AND FIELD STRENGTH

We employ the following phenomenological model of the dependence of the dielectric constant of a ferroelectric on the temperature and electric field strength[7,8]:

$$\varepsilon(\xi_s,T,E)=\varepsilon_{00}[(a(\xi_s,T,E)^{1/2}+\xi(\xi_s,E)^{2/3}+(a(\xi_s,T,E)^{1/2}-\xi(\xi_s,E)^{2/3}-\eta(T)]^{-1}, \tag{22}$$

where T is the temperature and E is the field strength.

The remaining components of formula (22) are given by

$$\eta(T)=\frac{T_F}{4T_C}\sqrt{1+\left(\frac{4T}{T_F}\right)^2}-1,$$

$$\xi(\xi_s,E)=\sqrt{\xi_s^2+\left(\frac{E}{E_N}\right)^2},$$

$$a(\xi_s,T,E)=\xi(\xi_s,E)^2+\eta(T)^3. \tag{23}$$

Here we have used the following model parameters: the parameter ε_{00} is defined in terms of the Curie–Weiss constant; T_C is the Curie temperature; T_F is the effective Debye temperature associated with those vibrations of the crystalline lattice that are responsible for the ferroelectric polarization, this parameter plays an important role in modeling the properties of incipient ferroelectrics such as strontium titanate ($SrTiO_3$) and potassium titanate ($KTaO_3$); E_N is the normalizing field, defining the dielectric nonlinearity of the material; ε_s is a parameter characterizing internal strains in the crystal and may be considered as a measure of imperfections in the crystal.

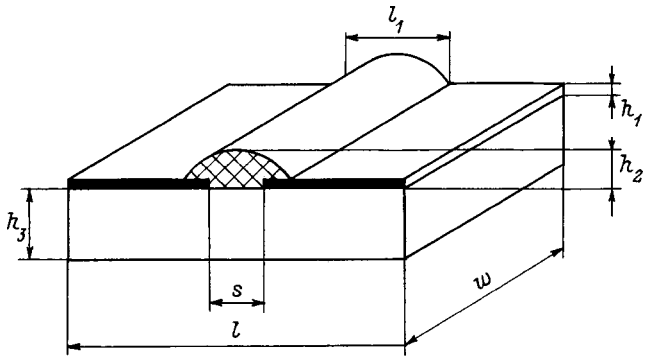

FIG. 6. Diagram of the planar capacitor investigated experimentally (Ref. 21).

Table II gives numerical values of the model parameters for $SrTiO_3$ films,[7,8] obtained by processing experimental data for a single crystal and for a series of thin-film sandwich capacitors.

Table II also lists the values of the Curie–Weiss constant C, which is the most stable characteristic of the ferroelectric state of the material.

COMPARISON WITH EXPERIMENT

To compare with experiment, we select out a study of a planar capacitor based on a $SrTiO_3$ film incorporated in a microstrip microwave resonator.[21] Figure 6 shows the design of the resonator with the incorporated planar capacitor. Dimensions of the design elements are listed in Table III.

The capacitance of the planar capacitor was measured from the shift of the resonant frequency of the resonator (10–12 GHz). Note that the dielectric constant of $SrTiO_3$ does not experience any noticeable dispersion up to frequencies on the order of 100 GHz. Applied to the given design, formula (21), obtained above for a planar capacitor of more general design, requires some correction.

1. The ferroelectric layer extends above the electrodes; therefore, its effective dielectric constant in the last term in formula (21) should be represented as

$$\varepsilon_2^*=\varepsilon_2-1. \tag{24}$$

2. If the capacitance of the capacitor in question is measured at low frequencies, then the total length of the electrodes $l=9$ mm should be taken into account when calculating the air capacitance. However, in measurements at microwave frequencies the electrodes form a resonator and

TABLE III. Dimensions of a microstrip resonator and the incorporated planar capacitor shown in Fig. 6.

	Thickness, μm	ε_r	Length, μm	Width w, μm
Electrodes	$h_1=0.3$	⋯	$l=9000$	3730
$SrTiO_3$ layer	$h_2=0.5$	$\varepsilon_2=\varepsilon(\xi_s,T,E)$	$l_1=100$	3730
$LaAlO_3$ layer	$h_3=430$	$\varepsilon_3=22$	$l=9000$	3730

their dimensions determine its unperturbed resonant frequency. In this case the air partial capacitance is formed by the capacitance of the gap in the microstrip line.[22] This capacitance can be described approximately as the capacitance per unit length of the short gap in a conducting screen:[22]

$$C=\varepsilon_0 w\frac{2}{\pi}\left[\ln\left(\frac{w}{s}\right)+1\right]. \tag{25}$$

Taking the above into account, we may rewrite formula (21) in the following form:

$$C=\varepsilon_0 w\cdot\left\{\frac{2}{\pi}\left[\ln\left(\frac{w}{s}\right)+1\right]+\frac{\varepsilon_3-1}{\pi}\ln\left(16\frac{h_3+h_2}{\pi s}\right)+\frac{\varepsilon_2-1}{s/h_2+(4/\pi)\ln 2\sqrt{1+2\varepsilon_2/\alpha h_2}}\right\}. \tag{26}$$

Note also that in the calculation of the capacitance of a planar capacitor we have ignored the thickness of the electrodes.

Reference 21 shows the dependence of the measured capacitance of a planar capacitor on the temperature and the applied voltage. We convert the applied voltage into bias field strength by simple formula

$$E=\frac{U}{s}. \tag{27}$$

This calculation is a rough approximation, since the real field distributio n in a planar capacitor is determined by the inhomogeneity of the boundary conditions and the nonlinearity of the medium.

We gather all the above-said into formula (26). We find the model parameters of the ferroelectric $\varepsilon_{00}, T_C, T_F, E_N$, and ξ_s that minimize the mean-square deviation of the calculated and measured capacitances. The results of this search are listed in Table IV.

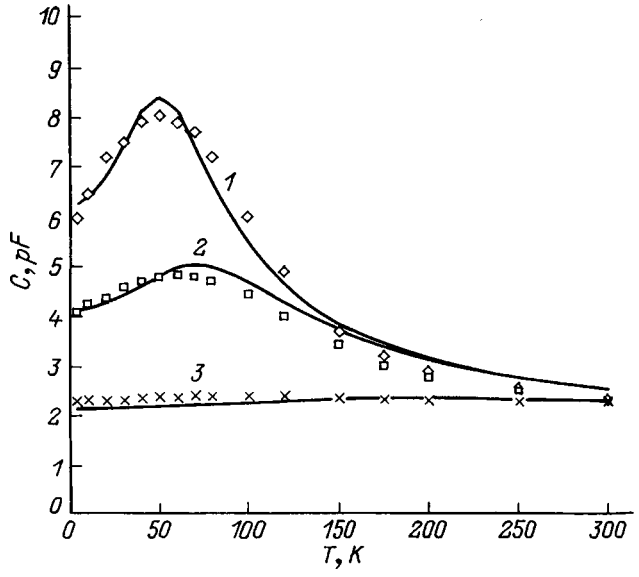

FIG. 7. Capacitance of a planar capacitor based on a $SrTiO_3$ film plotted as a function of the temperature for different values of the applied voltage: solid curves — calculation, points — experiment; Applied voltage: *1* — 0, *2* — 5, *3* — 50 V.

TABLE IV. Model parameters for the $SrTiO_3$ film entering into the makeup of the planar capacitor.

Model parameter	ε_{00}	T_C, K	T_F, K	E_N, kV/cm	ξ_s	$C=\varepsilon_{00}T_C$, K
Found value	1890	46.5	158	8	0.456	8.8×10^4

Figure 7 plots the experimental and the calculated dependence of the capacitance of a planar capacitor on the temperature and field strength. Comparison of the data listed in Tables II and IV shows that the parameters of the material used in Ref. 21 lie within the limits of variation associated with different technologies of manufacture of strontium titanate films.

CONCLUSION

The model considered allows one to calculate the capacitance of a layered planar structure containing a thin layer of ferroelectric material. The calculation is based on the partial capacitance method and conformal mapping. The partial capacitance method is based on the introduction of ''magnetic walls'' between the layers, i.e., on the assumption that the field lines are parallel to the boundaries of the media and do not extend beyond them. Applicability of the partial capacitance method is vouchsafed by fulfillment of the corresponding conditions on the geometrical dimensions of the structure.

The exact but cumbersome formulas given by conformal mapping can be simplified significantly by expanding in a small parameter. If the discussed conditions are satisfied, the error of using the simplified formulas to calculate the capacitance of planar capacitors does not exceed 1–2%. The previously expounded model description of the dependence of the dielectric constant of an incipient ferroelectric on the temperature and constant electric field strength combines nicely with the partial capacitance method. Taking the size effect into account leads to the appearance of an additional term in the model formula for the capacitance of a thin ferroelectric layer. Figure 7 shows the good agreement obtaining between the corresponding calculation and the experimental data for the dependence of a two-layer planar capacitor on the temperature and applied voltage.

[1] O. G. Vendik, L. T. Ter-Martirosyan, A. I. Dedyk *et al.*, Ferroelectrics **144**, No. 1–4, 33–43 (1993).
[2] O. G. Vendik, I. G. Mironenko, and L. T. Ter-Martirosyan, Microwaves RF **33**, No. 7, 67 (1994).
[3] S. S. Gevorgian, D. I. Kaparkov, and O. G. Vendik, Proc. IEE (London) **30**, 1236 (1994).
[4] A. T. Findikoglu, Q. X. Jia, I. H. Campbell *et al.*, Appl. Phys. Lett. **66**, 3674 (1995).
[5] Special Issue on the Microwave and Millimeter Wave Applications of High-Temperature Superconductivity, IEEE Trans. Microwave Theory Tech. **MTT-44**, 1193 (1996).
[6] *Ferroelectrics in Microwave Technology*, edited by O. G. Vendik [in Russian], Sov. Radio, Moscow, 1979.

[7] O. G. Vendik and S. P. Zubko, Zh. Tekh. Fiz. **67**, No. 3, 29 (1997) [Tech. Phys. **42**, 278 (1997)].
[8] O. G. Vendik, S. P. Zubko, and L. T. Ter-Martirosyan, Fiz. Tverd. Tela (St. Petersburg) **38**, 3654 (1996) [Phys. Solid State **38**, 1991 (1996)].
[9] V. I. Lavrik and V. N. Savel'ev, *Handbook on Conformal Transformations* [in Russian], Naukova Dumka, Kiev, 1970.
[10] É. S. Kochanov, Radiotekhnika **22**, No. 7, 82 (1967).
[11] É. S. Kochanov, Radiotekhnika **30**, No. 1, 92 (1975).
[12] K. R. Hoffman, *Handbook of Microwave Integrated Circuits* (Artech House, Norwood, Massachusetts, 1987).
[13] S. S. Bedair and I. Wolf, IEEE Trans. Microwave Theory Tech. **MTT-40**, 41 (1992).
[14] S. S. Gevorgian, L. J. P. Linnér, and E. L. Kollberg, IEEE Trans. Microwave Theory Tech. **MTT-43**, 772 (1995).
[15] S. S. Gevorgian, T. Martinsson, and L. J. P. Linnér *et al.*, IEEE Trans. Microwave Theory Tech. **MTT-44**, 896 (1996).
[16] E. Chen and S. Y. Chou, IEEE Trans. Microwave Theory Tech. **MTT-45**,939 (1997).
[17] A. N. Deleniv, Zh. Tekh. Fiz. **69**(4), 8 (1999) [Tech. Phys. **44**, 356 (1999)].
[18] K. Gupta, R. Garg, and R. Chadkha, *Computer-Aided Design of Microwave Circuits* (Artech House, Norwood, Massachusetts, 1981).
[19] O. G. Vendik, I. G. Mirinenko, and L. T. Ter-Martirosyan, Fiz. Tverd. Tela (Leningrad) **26**, 3094 (1984) [Sov. Phys. Solid State **26**, 1864 (1984)].
[20] O. G. Vendik and L. T. Ter-Martirosyan, Fiz. Tverd. Tela (St. Petersburg) **36**, 3343 (1994) [Phys. Solid State **36**, 1778 (1994)].
[21] D. Galt, C. Price, J. A. Beall *et al.*, Appl. Phys. Lett. **63**, 3078 (1993).
[22] I. Kasa, *Microwave Integrated Circuits* (Elsevier, New York, 1993).

Translated by Paul F. Schippnick

On the question of the error in the partial capacitance method

A. N. Deleniv

St. Petersburg State Electrical Engineering University, 197376 St. Petersburg, Russia
(Submitted December 24, 1997)
Zh. Tekh. Fiz. **69**, 8–13 (April 1999)

An analytical method is proposed for calculating the capacitance of a planar capacitor constructed in the form of a two-layer structure. The method is used to calculate the capacitance of a planar capacitor containing a ferroelectric film. A number of calculations were performed to check the accuracy of the partial capacitance method. The calculations made it possible to determine the limits of accuracy of the partial capacitance method for a permissible calculational error of $\pm 3\%$. © *1999 American Institute of Physics.* [S1063-7842(99)00204-4]

INTRODUCTION

Ferroelectrics are very attractive materials on which to base controllable

microwave devices. The large dielectric constant of ferroelectrics allows them to be used as elements of microwave circuits in the form of thin films with thickness of the order of 1 μm (Ref. 1). Of substantial importance is the possibility of reliable calculation of the parameters of planar structures containing a thin film with a large dielectric constant. The basis of the calculation in the quasistatic approximation is the partial capacitance method (PCM). The partial capacitance method was first used by Kochanov.[2,3] The use of this method to calculate planar structures requires an estimate of their accuracy. The first such estimate was made by Kochanov himself[3] by comparison against a calculation performed on an analog computer: the error did not exceed 1–2%. Later the partial capacitance method was included in a handbook[4] and recommended for calculations as *a priori* reliable. The partial capacitance method is also used to determine the dielectric characteristics of ferroelectric films,[5] where a planar capacitor containing a ferroelectric film is used as a metrological element, and all the calculations are made using the partial capacitance method.

The main goal of the present work is to verify the reliability of the partial capacitance method in relation to calculations of the capacitances

of planar structures similar to the one shown in Fig. 1. A method is also proposed combining methods of conformal mapping and mirror images and allowing one with good accuracy to calculate the capacitances of planar structures. To simplify the calculation, the capacitor electrodes are assumed to be semi-infinite, which is not a fundamental point.

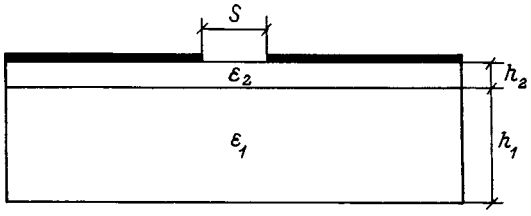

FIG. 1. Two-layer planar capacitor with a ferroelectric film (original structure).

DESCRIPTION OF THE METHOD

Let us turn our attention to the structure shown in Fig. 1. Thanks to the symmetry of the structure, an electric wall can be situated in the symmetry plane: Fig. 2a. We will use the method of conformal mapping.[6] In order to transform the original plane into the inner region of a plane-parallel capacitor, two successive mappings are needed. The first mapping function

$$W=\sin\left[\frac{\pi T}{2(h_1+h_2)}\right] \tag{1}$$

maps a semi-infinite strip of the T plane onto a quadrant of the W plane. For brevity, we do not display the corresponding graphical constructions. For the second mapping we use the Christoffel–Schwarz transformation

$$L=\int_0^w \frac{dw}{\sqrt{(1-w^2)(1-k^2w^2)}} \tag{2}$$

with modulus

$$k=\left[\cosh\left(\frac{s\pi}{4(h_1+h_2)}\right)\right]^{-1}.$$

The transformation maps the indicated quadrant of the W plane onto the inner region of a planar capacitor in the L plane (Fig. 2b). Here $K(k)$ is the elliptic integral of the first kind. The numbers in Fig. 2 label equivalent points of the original and transformed structures. The boundary between the dielectrics is transformed into a curve in the L plane, the form and position of which determine the capacitance of the investigated structure. Since the coordinates of the curve are related with one another implicitly (through the coordinates x,y of the T plane), for the calculations that are to follow it is useful to describe this curve in the coordinates w,f (Fig. 2). We represent the function $w(f)$ describing the boundary of the dielectrics in the form of a Fourier sine series expansion on an interval of length T, where T is the distance between the points *1–4* in Fig. 2,

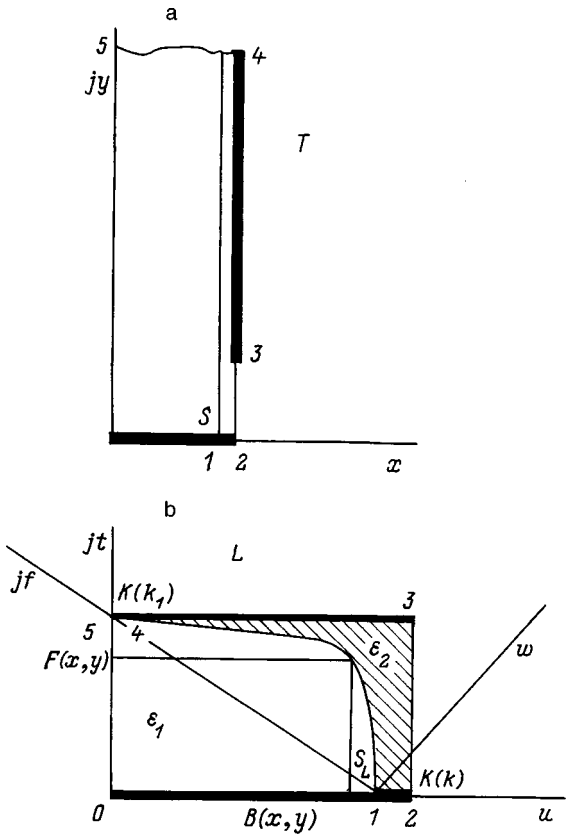

FIG. 2. Transformation of the original structure: a — capacitor divided in half by an ideal conducting wall, b — result of conformal mapping onto the plane $L \subset w+jt$. The surface separating the two regions with different dielectric constant in the resulting planar capacitor is indicated.

$$w(f)=\sum_{n=1}^{M} D_n \sin(n2\pi f/T), \tag{3}$$

where

$$T=2\sqrt{S_L^2+K(k_1)^2},$$

$$S_L=\int_0^{\sin(0.5h_1\pi/(h_1+h_2))} \frac{dw}{\sqrt{(1-w^2)(1-k^2w^2)}},$$

$$k_1=\sqrt{1-k^2}.$$

Here $K(k)$ is the elliptic integral of the first kind. The expansion coefficients D_m are found by numerical integration of relations obtained by conformal mapping of the boundary of the dielectrics.

The electrostatic field, being the solution for the given closed region, should satisfy boundary conditions and the homogeneous equation

$$\mathrm{div}\{\varepsilon(f,w)\mathrm{grad}[\varphi(f,w)]\}=0.$$

Using the formulas of vector analysis, we can rewrite this equation in the following form:

$$\mathrm{grad}[\varepsilon(f,w)]\mathrm{grad}[\varphi(f,w)]+\varepsilon(f,w)\Delta\varphi(f,w)=0. \tag{4}$$

The first term in Eq. (4) vanishes for a homogeneous medium [Eq. (4) goes over to the Laplace equation] or leads to the appearance of a delta function at the boundary between the two media for the case under consideration. Formally, it can be assumed that Eq. (4) is the Poisson equation, where the first term describes a charge distributed over an infinitely thin layer on the boundary between the media. Taking all the above-said into account, it may be recapitulated that introducing a surface charge has the same effect on the field of a planar capacitor as the boundary between the dielectric media. Thus, the interesting electrostatic

field will be represented as a superposition of two components: an external uniform field defined by a voltage of 1 V applied to the electrodes of the capacitor

$$E_{\mathrm{ext}}(u,t)=\frac{1}{jK(k_1)} \tag{5}$$

and the field of the surface charges $E_\sigma(u,t)$. Consequently, the solution of the problem reduces to finding the surface charge $\sigma(f)$ and the corresponding field $E_\sigma(u,t)$.

The surface charge distribution $\sigma(t)$ should satisfy boundary conditions for the normal field components at the boundary between the dielectrics

$$\frac{E^n_{\mathrm{ext}}(f)+E^n_\sigma(f)+\sigma(f)/2\varepsilon_0}{E^n_{\mathrm{ext}}(f)+E^n_\sigma(f)-\sigma(f)/2\varepsilon_0}=\frac{\varepsilon_2}{\varepsilon_1}, \tag{6}$$

where ε_0 is the dielectric constant of the vacuum; ε_1 and ε_2 are the relative dielectric constants of the substrate and the ferroelectric layer, respectively; $E^n_{\mathrm{ext}}(f)$ is the normal component of the external field at the boundary; $E^n_\sigma(f)$ is the normal component of the field due to the charge distribution $\sigma(f)$ on the boundary; $\sigma(f)/2\varepsilon_0$ is a component producing a discontinuity in the normal component in the electric field.

FIELD DISTRIBUTION OF A LINEAR CHARGE AT THE BOUNDARY OF TWO DIELECTRICS FOR GIVEN BOUNDARY CONDITIONS

We distinguish a strip of width $d\lambda$ on the boundary of the two media, perpendicular to the plane of the figure (Fig. 3). We treat this strip as a uniformly charged thread with linear charge density $d\tau=\sigma(f_a)d\lambda$ [C/m] at the point with coordinates f_a, $w(f_a)$. To solve the stated problem, it is necessary to find the Green's function for the region bounded by two electric and two magnetic walls in the L plane. In order to make use of a ready solution, we transform the original

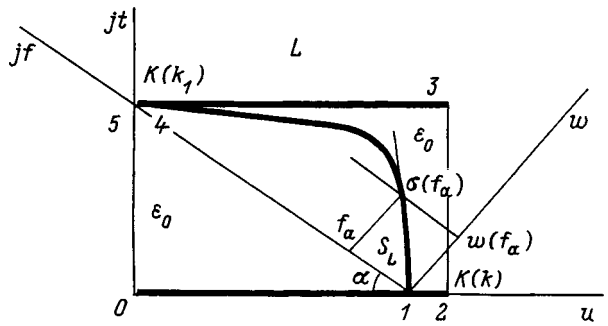

FIG. 3. The same structure as in Fig. 2, with the difference that the dielectric filler is uniform and an electric charge is found on the interface, with surface density $\sigma(f)$.

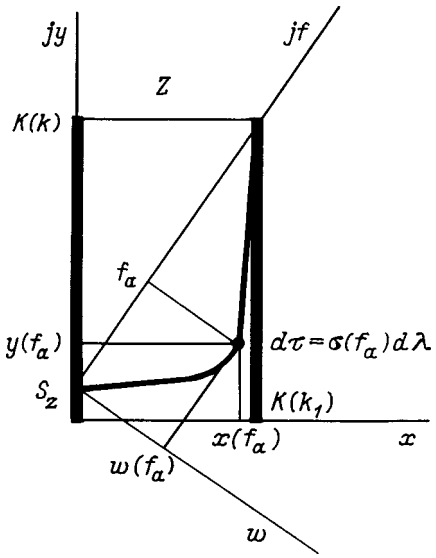

FIG. 4. The structure under consideration is transformed to facilitate finding the electrostatic field induced by a charged thread running along the line with coordinates $x(f_a)$, $y(f_a)$.

coordinate plane (the L plane) according to the rule $Z=(K(k)-L)\exp(j\pi/2)$ (Fig. 4). The calculation of the potential and the field of a linear charge located between conducting planes was considered in Ref. 7. The solution for the potential created by an isolated charged thread between conducting planes is written in the form of the complex function

$$d\omega(Z,f_a)=\ln\left[\sin\pi\left(\frac{Z+\overline{Z}(f_a)}{2K(k_1)}\right)\Big/\sin\pi\left(\frac{Z-Z(f_a)}{2K(k_1)}\right)\right]\times\frac{\sigma(f_a)d\lambda}{2\pi},\tag{7}$$

$Z(f)=x(f)+jy(f)$; $\overline{Z}(f)=x(f)-jy(f)$; and Z is the complex number formed by the coordinates of the observation point.

The complex function $Z(f)$ establishes a correspondence between the curve $w(f)$ (Fig. 4) and the coordinates x,y in the $Z(f)=x(f)+jy(f)$ plane. The solution for the case of interest to us is formed by an infinite series of pairs of threads, in each of which the threads are separated from one another by the same distance and equal in the sign of their charges and are situated symmetrically with respect to the magnetic walls. Thus, for the potential created by an isolated charged thread between conducting planes and magnetic walls, we can write

$$d\psi(Z,f_a)=\sum_{n=0}^{\infty}\ln\left[\frac{\sin\pi([Z-j2Kn+\overline{Z}(f_a)]/2K_1)\times\sin\pi([Z-j2K(n+1)+Z(f_a)]/2K_1)}{\sin\pi([Z-j2Kn-Z(f_a)]/2K_1)\times\sin\pi([Z-j2K(n+1)-\overline{Z}(f_a)]/2K_1)}\right.$$
$$\left.\times\frac{\sin\pi([Z+j2Kn+Z(f_a)]/2K_1)\times\sin\pi([Z+j2K(n+1)+\overline{Z}(f_a)]/2K_1)}{\sin\pi([Z+j2Kn-\overline{Z}(f_a)]/2K_1)\times\sin\pi([Z+j2K(n+1)-Z(f_a)]/2K_1)}\right]\frac{\sigma(f_a)d\lambda}{2\pi}.\tag{8}$$

Since the series is rapidly converging, it is sufficient to calculate 5–6 terms. For convenience in what is to follow, we rewrite expression (8) in the form

$$d\psi(Z,f_a)=D(Z,f_a)\,\sigma(f_a)\,d\lambda.\tag{9}$$

Expression (9) gives the electric potential created by a charged thread.

Now it is possible to find the field created by a charged thread

$$dE(Z,f_a)=\left|\frac{dD(Z,f_a)}{dZ}\right|\exp\left(-j\arg\left(\frac{dD(Z,f_a)}{dZ}\right)\right)\sigma(f_a)d\lambda.\tag{10}$$

DETERMINATION OF AN UNKNOWN SURFACE CHARGE DISTRIBUTION $\sigma(f)$

Condition (6) can be rewritten in the form

$$H\sigma(f)+E^n_{\text{ext}}(f)+E^n_{\sigma}(f)=0,\tag{11}$$

$$H=\frac{\varepsilon_2+\varepsilon_1}{2\varepsilon_0(\varepsilon_1-\varepsilon_2)}.\tag{12}$$

We represent all three terms of Eq. (11) as harmonic Fourier cosine series with period T

$$E^n_{\text{ext}}(f)=\sum_{n=0}^{m}C_n\cos\left(\frac{2\pi nf}{T}\right).\tag{13}$$

The coefficients of the series can be calculated directly since the equation of the curve describing the boundary of dielectrics is known and, consequently, its slope is also known. The function $E^n_{\sigma}(f)$ is related to the unknown charge distribution $\sigma(f)$ by the following integral operation:

$$E^n_{\sigma}(f)=\int_0^{T/2}\sigma(f')\,\text{Re}\left(\frac{dD(Z(f),f')}{dZ}\exp(j\beta(f))\right)\times\frac{df'}{\cos(\vartheta(f'))},\tag{14}$$

where $\vartheta(f)$ is the angle defining the slope of the tangent to the boundary in the $f-w$ coordinate plane; $\beta(f)=\vartheta(f)+\alpha$, where α is the angle between the coordinate axes f,w and t,u (Fig. 3).

We represent expression (14) as an integral transform with kernel $K(f,f')$

$$K(f,f')=\mathrm{Re}\left(\frac{dD(Z(f),f')}{dZ}\,\frac{\exp(j\beta(f))}{\cos(\vartheta(f'))}\right). \tag{15}$$

In the description of the kernel $K(f,f')$ it is necessary to use expansion (3) defining the coordinates of the boundary f_1, $w(f_a)$.

Taking into account that the surface charge is represented in the form

$$\sigma(f)=\sum_{n=0}^{m} A_n\cos\left(\frac{2\pi nf}{T}\right), \tag{16}$$

we obtain

$$E_\sigma^n(f)=\sum_{n=0}^{m} A_n\int_0^{T/2}\cos\left(\frac{2n\pi f'}{T}\right)K(f,f')\,df'. \tag{17}$$

We also represent $E_\sigma^n(f)$ as a series

$$E_\sigma^n(f)=\sum_{k=0}^{m} B_k\cos\left(\frac{2\pi kf}{T}\right) \tag{18}$$

and write down an expression for the coefficients of this series B_k

$$B_k=\sum_{n=0}^{m} A_nB_{k,n}, \tag{19}$$

where

$$B_{k,n}=\left[\frac{2}{T}\int_0^{T/2}\cos\left(\frac{2\pi kf}{T}\right)\int_0^{T/2}\cos\left(\frac{2\pi nf'}{T}\right)K(f,f')df'df\right]; \tag{20}$$

the $B_{k,n}$ are coefficients of the matrix representing the operator transforming from the surface charge distribution (16) to the field strength (18).

Now Eq. (11) can be rewritten as a system of linear equations in the coefficients A_n

$$\begin{bmatrix} H+B_{0,0} & B_{0,1} & \dots & B_{0,m} \\ B_{1,0} & H+B_{1,1} & \dots & B_{1,m} \\ \dots & \dots & \dots & \dots \\ B_{m,0} & B_{m,1} & \dots & H+B_{m,m} \end{bmatrix}\begin{bmatrix} A_0 \\ A_1 \\ \dots \\ A_m \end{bmatrix}=\begin{bmatrix} -C_0 \\ -C_1 \\ \dots \\ -C_m \end{bmatrix}. \tag{21}$$

Finding the unknown distribution $\sigma(f)$ reduces to finding the coefficients A_n. Toward this end, it is necessary to calculate the matrix B. Having calculated the matrix B with the aid of Eq. (20), after finding the coefficients A_n and invoking relations (13), (16), and (18) we find all the field components configuring the induction flux through the cross section of the capacitor shown in Fig. 4. As a result, we obtain the following formula for the capacitance:

TABLE I.

	$h_2=10\ \mu$m, $s=200\ \mu$m			$h_2=10\ \mu$m, $s=100\ \mu$m		
ε_2	C	$C^{(p)}$	ΔC, %	C	$C^{(p)}$	ΔC, %
300	0.397	0.391	1.60	6.642	6.556	1.30
1000	0.981	0.984	−0.33	1.790	1.795	−0.28
3000	2.648	2.680	−1.20	5.004	5.049	−0.90
5000	4.315	4.376	−1.40	8.218	8.303	−1.00
10 000	8.483	8.617	−1.60	16.250	16.440	−1.20
	$h_2=5\ \mu$m, $s=80\ \mu$m			$h_2=5\ \mu$m, $s=20\ \mu$m		
300	0.506	0.500	1.20	1.347	1.326	1.56
1000	1.225	1.234	−0.74	3.883	3.864	0.49
3000	3.279	3.332	−1.60	11.130	11.120	0.09
5000	5.333	5.43	−1.80	18.37	18.37	0.00
10 000	10.470	10.67	−1.90	36.5	36.51	−0.03
	$h_2=1\ \mu$m, $s=20\ \mu$m			$h_2=1\ \mu$m, $s=5\ \mu$m		
300	0.524	0.519	0.82	1.243	1.224	1.53
1000	1.097	1.113	−1.46	3.341	3.332	0.28
3000	2.736	2.809	−2.67	9.334	9.352	−0.19
5000	4.374	4.504	−2.97	15.330	15.370	−0.26
10 000	8.470	8.744	−3.20	30.320	30.420	−0.33

$$C=\varepsilon_0\varepsilon_1\int_0^{T/2}\left(\frac{E_{\mathrm{ext}}^n(f)+E_\sigma^n(f)-\sigma(f)/2\varepsilon_0}{\cos(\vartheta(f))}\right)df+\varepsilon_0\varepsilon_1 \times\left[\int_0^{S_z}\int_0^{T/2}\frac{dD(Z,f)}{dZ}\,\frac{\sigma(f)}{\cos(\vartheta(f))}\,df\,dZ+\frac{K(k)-S_z}{K(k_1)}\right]. \tag{22}$$

The quantity S_z entering into the formula is defined as the intersection point of the boundary of the dielectrics and the y axis (Fig. 4).

DISCUSSION OF THE ACCURACY OF THE CALCULATION

Errors in the calculations using the partial capacitance method are due mainly to one source—the representation of the function $w(f)$ describing the boundary between the dielectrics in the form of a finite series. The second source of error—the finite number of terms in expression (8)—is unimportant, and therefore will not be considered further. Using a different number of base functions in the calculations, that is to say, a different size of the matrix B assigned by Eq. (20), it is possible to estimate the error. It was found that for the size of the matrix B equal to 15–16 the limiting values of the capacitance of the structure differ from the calculated values by not more than 0.2%.

THE PARTIAL CAPACITANCE METHOD

The application of the partial capacitance method to the structure under consideration was discussed in detail in Ref. 5, and therefore only calculational formulas are given below

$$C^{(p)}=C_1^{(p)}+C_2^{(p)}, \tag{23}$$

where $C_1^{(p)}$ and $C_2^{(p)}$ are the capacitance of the substrate and of the ferroelectric, respectively.

In the calculation I used the simplified formula[6]

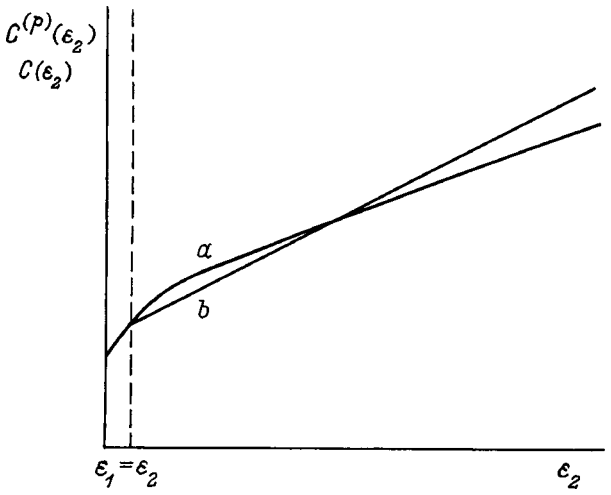

FIG. 5. Qualitative shape of the dependence of the capacitance of a two-layer structure as a function of the dielectric constant ε_2: a — the proposed method, b — the partial capacitance method.

$$C^{(p)}=\varepsilon_0\left(\varepsilon_1\frac{1}{\pi}\ln\left(\frac{16(h_1+h_2)}{\pi s}\right)+\frac{\varepsilon_2-\varepsilon_1}{s/h_2+(4/\pi)\ln(2)}\right). \tag{24}$$

Note that the partial capacitance method is inapplicable in the case when $\varepsilon_2<\varepsilon_1$, i.e., when the second term in Eq. (24) becomes negative.

COMPARISON AND DISCUSSION OF RESULTS

To check the accuracy of the partial capacitance method, I performed a number of calculations for different geometries over a wide range of variation of $\varepsilon_2/\varepsilon_1$. Results of the calculation s are shown in Table I. All calculations were made for a substrate with $\varepsilon_1=10$ and thickness $h=500\,\mu$m; capacitances are indicated in nanofarads (nF). Figure 5 gives a qualitative picture of the dependence of $C=f(\varepsilon_2)$ obtained by the proposed method (curve *a*) and $C^{(p)}=f(\varepsilon_2)$ obtained using the partial capacitance method (curve *b*). The second curve is constrained by the condition $\varepsilon_2>\varepsilon_1$. It is obvious that there exist two sources of error in the partial capacitance method at small and large values of ε_2. The partial capacitance method underestimates the capacitance at low values of ε_2 and overestimates it at high values of ε_2.

Figure 6 shows a contour plot of equal-error curves in the coordinates *s*, ε_2. A typical value of the error for geometries of practical interest is 1–2%. However, it should be borne in mind that at large values of the ratios $s/h_2(\geqslant 100)$ and $h_1/h_2\,(\geqslant 500)$ the error can significantly exceed its expected value $(\geqslant 5)$. It should also be noted that the extraordinary

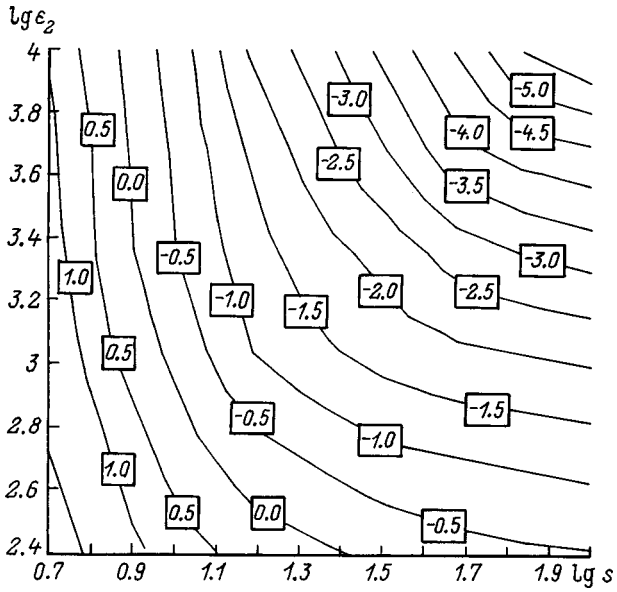

FIG. 6. Different error curves in the coordinates *s*, ε_2 (*s* in μm).

simplicity and good accuracy of the partial capacitance method makes it a reliable tool for practical calculations within the limits of the above-mentioned restrictions.

I am indebted to Prof. O. G. Vendik for posing the problem and for useful discussions; I am also very grateful to S. P. Zubko for help with the text.

[1] *Ferroelectrics in Microwave Technology* [in Russian], edited by O. G. Vendik [in Russian], Sov. Radio, Moscow (1979), 272 pp.
[2] É. S. Kochanov, Radiotekhnika **22**(7), 82 (1967).
[3] É. S. Kochanov, Radiotekhnika **30**(1), 92 (1975).
[4] K. R. Hoffman, *Handbook of Microwave Integrated Circuits* (Artech House, Norwood, Massachusetts, 1987).
[5] O. G. Vendik, S. P. Zubko, and M. A. Nikol'skiĭ, Zh. Tekh. Fiz. **69**(4), 1 (1999) [Tech. Phys. **41**, 349 (1999)].
[6] V. I. Lavrik and V. N. Savel'ev, *Handbook on Conformal Transformations* [in Russian], Naukova Dumka, Kiev (1970), 252 pp.
[7] K. J. Binns and P. J. Lawrenson, *Analysis and Calculation of Electric and Magnetic Fields* (Pergamon Press, Oxford, 1963; Énergiya, Moscow, 1970).

Translated by Paul F. Schippnick

Effect of molecular nitrogen on the electron mobility in a mixture of argon and optically excited sodium vapor

N. A. Gorbunov and A. S. Mel'nikov

Scientific-Research Institute of Physics, St. Petersburg State University, 198904 St. Petersburg, Russia
(Submitted November 10, 1997)
Zh. Tekh. Fiz. **69**, 14–19 (April 1999)

A parametric study of the electron energy distribution function (EEDF) and the electron mobility in the mixture Na + Ar + N_2 is carried out. An analysis is made of the conditions that obtain in a photoplasma when the detachment of the mean electron energy from the neutral gas temperature is due to superelastic collisions (collisions of the second kind) with excited sodium atoms. The case of low ionization of the medium at low vibrational temperatures of the ground state of the nitrogen molecules is considered. To find the EEDF a numerical solution of the Boltzmann transport equation is carried out. It is found that in the indicated mixture the presence of nitrogen leads to a depletion of the EEDF in the region of efficient vibrational excitation of the molecules and promotes the formation of inversion in the EEDF $\partial f(\varepsilon)/\partial\varepsilon>0$ in the energy range corresponding to the Ramsauer minimum in the cross section of elastic collisions of electrons with the argon atoms. It is shown that the nonequilibrium character of the EEDF leads to a complicated dependence of the electron mobility on the partial ratios of the components of the mixture, the degree of ionization of the medium, and the population of the resonantly excited sodium atoms. © *1999 American Institute of Physics.* [S1063-7842(99)00304-9]

INTRODUCTION

The electron transport parameters in a weakly ionized gas depend on the electron energy distribution function (EEDF), which, as a rule, is nonequilibrium under actual experimental conditions. In contrast to a plasma of atomic gases, in a molecular plasma excitation of rotational and vibrational levels is possible. For a number of gases, such as nitrogen and carbon dioxide, the excitation cross sections of the vibrational levels of the electron ground state have a number of maxima grouped together in a narrow (on the order of 1 eV) energy interval. This is because vibrational excitation takes place via formation of an unstable negative ion. The indicated nonmonotonic dependences give rise to a complicated form of the EEDF in a molecular plasma.

Numerous works have been dedicated to the calculation of the EEDF in a molecular-nitrogen plasma under gas-discharge conditions (see Refs. 1–3 and the literature cited therein). It has been shown that processes of excitation of vibrational levels play a defining role in the electron energy balance over a wide range of experimental conditions. It has been established that the form of the EEDF in the region of efficient vibrationa l excitation and de-excitation depends on the distribution of the molecules over the vibrational states.[4–6] Vibrationally excited molecules have been found to have an effect on the electron transport parameters in a molecular plasma.[7] It has been shown[8] that in a mixture of the above-enumerated molecular gases and heavy inert gases the appearance of a negative differential conductivity is possible in the plasma.

The distribution of the molecules over vibrational levels has its strongest effect on the form of the EEDF in non-self-sustaining discharges under the weak influence of a longitudinal electric field when heating of the electron gas is due to collisions with the excited molecules. The EEDF for the afterglow of a discharge in pure nitrogen was calculated[9,10] for various degrees of ionization of the medium and different vibrational temperatures of the electron ground state. The influence of the occupancie s of the electron-excited states on the high-energy part of the EEDF has been modeled. The energy distribution in a nitrogen afterglow for the vibrational temperature of the ground state $T_v=0.3$ eV was measured in Ref. 11. This same reference gives an analytical calculation of the EEDF based on representing the inelastic electron–molecule collision integral in the Fokker–Planck approximation. Detailed numerical calculations for the conditions of the experiment in Ref. 11 were performed in Refs. 12 and 13. Reference 14 presents the results of measurements of the EEDF in an electron-beam low-temperature nitrogen discharge under conditions in which the vibrational temperature was near the gas temperature and equal roughly to 0.03 eV.

An analysis of the results of Refs. 9–14 shows that the EEDF in non-self-sustaining nitrogen discharges (discharge afterglow, beam discharge, photoplasma) in the energy range from $\varepsilon_1=1.5$ eV to $\varepsilon_2=3.6$ eV corresponding to the range of efficient vibrational excitation depends substantially on the distribution of the molecules over vibrational levels of the electron ground state. In the case when the vibrational temperature T_v is near the gas temperature T_a in the indicated energy range, a depletion of the EEDF due to excitation of vibrational levels is observed. With growth of the vibrational temperature the number of electrons in this energy interval increases due to superelastic collisions (collisions of the sec-

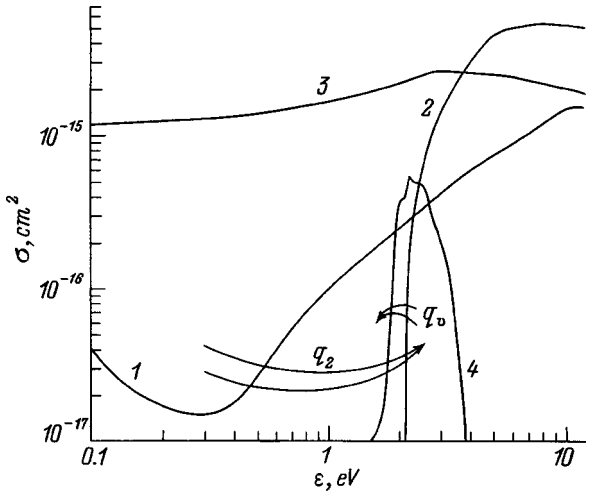

FIG. 1. Collision cross sections of the main processes forming the EEDF in a Na + Ar + N_2 mixture: *1* —transport cross section of elastic collisions of electrons with the argon atoms, *2* — excitation cross section of $3S-3P$ resonance levels, *3* — cross section of superelastic collisions with excited sodium atoms, *4* — total excitation cross section of the eight vibrational levels $X^1\Sigma_g^+$ of the nitrogen molecule.

ond kind) with vibrationally excited molecules. If $T_v \gg T_a$, then a quasi-equilibrium distribution of the electrons is formed in the energy interval from ε_1 to ε_2, whose temperature is near T_v. The form of the EEDF for energies $\varepsilon < \varepsilon_1$ depends substantially on the degree of ionization of the plasma.

In our previous works[15–18] we considered the mechanism of formation of the EEDF under conditions of a non-self-sustaining discharge in two-component mixtures of sodium and inert gases. We showed that in such mixtures the nonequilibrium character of the distribution function substantially alters the electron mobility.[16] The influence of the appearance of inversion on the EEDF explains the effect observed earlier[19] of low conductivity of a photoplasma in a Na + Ar mixture.[17] In the present work, by inverted distribution we mean an EEDF having positive derivative $[\partial f(\varepsilon)/\partial\varepsilon > 0]$ in some energy interval. In Refs. 15 and 18 we established the existence of a critical population of the resonantly excited sodium atoms, above which an inversion is created in the EEDF such that a negative electron mobility effect is possible in the weakly ionized plasma in a mixture with heavy inert gases. However, in a two-component mixture this effect is possible only in a narrow range of conditions which are difficult to achieve experimentally. To enhance the negative mobility effect, we propose to use an admixture of nitrogen, which has a pronounced maximum in the excitation cross section of the vibrational levels, which makes it possible to alter the form of the EEDF in a controlled manner.

A qualitative explanation of the enhancement of inversion in the EEDF in a three-component mixture is illustrated by Fig. 1, which plots cross sections of the main processes forming the electron energy distribution.

The argon in the example mixture defines the necessary growth of the transport collision frequency of the electrons at energies exceeding the Ramsauer minimum. To obtain inversion in the EEDF it is necessary to form a fast electron source in the growing part of the cross sections of the elastic collisions of the electrons with the argon atoms. Under the conditions in question, fast electrons are formed as a result of superelast ic collisions with excited sodium atoms (the flux q_2 in Fig. 1).

In a two-component mixture the efficiency of the source is reduced as a result of inelastic collisions (collisions of the first kind) with the sodium atoms. Therefore, it is of interest to consider a three-component mixture, where excitation of vibrational levels of the nitrogen molecules (the flux q_v in Fig. 1) leads to a decrease in the frequency of inelastic collisions with Na. The choice of nitrogen as the third component of the mixture is motivated by the fact that the maximum of the excitation cross section of the vibrational levels of N_2 is located in the near-threshold region of sodium excitation. The high efficiency of nitrogen in forming an inverted EEDF is due to the fact that fast electrons leave the zone of excitation of resonant sodium levels with small energy losses compared to the excitation threshold.

The aim of the present work is to analyze the mechanism of formation of a nonequilibrium EEDF in non-self-sustaining discharges for three-component mixtures consisting of argon, sodium, and molecular nitrogen. Manifestations of the nonlocal character of the EEDF in a low-pressure molecular plasma are considered in Refs. 20–22.

THEORETICAL MODEL

To calculate the EEDF, we used the Boltzmann equation, which in the steady-state approximation has the form

$$\frac{d}{d\varepsilon}\left[D(\varepsilon)\frac{df(\varepsilon)}{d\varepsilon}+V(\varepsilon)f(\varepsilon)\right]=C(\varepsilon), \tag{1}$$

where $D(\varepsilon)$ is the diffusion coefficient in energy space

$$D(\varepsilon)=\varepsilon^{3/2}\left\{\left(\delta\nu_{el}(\varepsilon)+\frac{8B_e}{\varepsilon}\nu_{\text{curl}}(\varepsilon)\right)T_a+\frac{4}{3}\nu_{ee}(\varepsilon)\right.$$
$$\left.\times\left[\int_0^\varepsilon f(\varepsilon')\varepsilon'^{3/2}d\varepsilon'+\varepsilon^{3/2}\int_\varepsilon^\infty f(\varepsilon')d\varepsilon'\right]\right\}, \tag{2}$$

$$\delta\nu_{el}(\varepsilon)=v\sum_j\frac{2m}{M_j}n_j\sigma_{el}^j(\varepsilon) \tag{3}$$

is the effective frequency of energy losses in the elastic collisions, summed over all components of the mixture with concentrations n_j and cross sections $\sigma_{el}^j(\varepsilon)$; v and m are the velocity of the electron and its mass, respectively; M_j is the mass of the atoms (molecules) of species j; $B_e=2.47\times10^{-4}$ eV is the rotational constant of the nitrogen molecule; $\nu_{\text{curl}}(\varepsilon)$ is the excitation frequency of the rotational levels; $\nu_{ee}(\varepsilon)$ is the electron–electron collision frequency[12];

$$V(\varepsilon)=\varepsilon^{3/2}\left[\delta\nu_{el}(\varepsilon)+\frac{8B_e}{\varepsilon}\nu_{\text{curl}}(\varepsilon)+2\nu_{ee}(\varepsilon)\int_0^\varepsilon f(\varepsilon')\sqrt{\varepsilon'}d\varepsilon'\right] \tag{4}$$

is the dynamic friction coefficient in energy space;

$$C(\varepsilon)=\sum_{i,k}\left\{\varepsilon\sigma_{ik}(\varepsilon)\left[n_i f(\varepsilon)-\frac{g_i}{g_k}n_k f(\varepsilon-E_{ik})\right]-(\varepsilon+E_{ik})\sigma_{ik}(\varepsilon+E_{ik})\times\left[n_i f(\varepsilon+E_{ik})-\frac{g_i}{g_k}n_k f(\varepsilon)\right]\right\} \quad (5)$$

is the inelastic collision integral; $\sigma_{ik}(\varepsilon)$ is the excitation cross section of the kth level with population n_k and statistical weight g_k from the ith state with population n_i and statistical weight g_i; and E_{ik} is the energy gap between the levels.

The normalization of the EEDF is given by

$$\int_0^\infty f(\varepsilon)\sqrt{\varepsilon}\,d\varepsilon=1. \quad (6)$$

The electron mobility is expressed in terms of the isotropic part of the EEDF as follows:

$$\mu=-\frac{2e}{3m}\int_0^\infty\frac{\partial f(\varepsilon)}{\partial\varepsilon}\frac{\varepsilon^{3/2}}{\nu_{el}(\varepsilon)}d\varepsilon, \quad (7)$$

where e is the charge of the electron.

Equation (1) was solved numerically by the trial-and-error method.[15]

Elastic cross sections of the electrons with sodium and inert gases were taken from Refs. 23 and 24, and the Na($3S-3P_{1/2,3/2}$) excitation cross section was taken from Ref. 25. The effective excitation cross section of the rotational levels of the nitrogen molecule was taken from Ref. 26, the excitation cross sections of the vibrational levels were taken N_2 from Ref. 27, the energies of the vibrational levels corresponded to Ref. 28, and the transport cross section of the elastic collisions for nitrogen is given in Ref. 24. The cross sections of excitation from the ground state of N_2 to the higher-lying electron states were chosen in an analogous way as in Refs. 11 and 29.

We take the distribution of the molecules over vibrational levels of the ground state to be Boltzmannian with temperature T_v, which we take to be equal to the temperature of the atoms T_a. The gas temperature in the calculations was varied within the limits 450–600 K, which corresponds to sodium densities in the range $10^{12}-10^{15}\,\mathrm{cm}^{-3}$. For a characteristic size of the plasma greater than 1 cm (under the conditions considered below) for these densities the local approximation is valid in the calculation of the EEDF. We characterize the total population N* of the resonant states of sodium $3P_{1/2}$ and $3P_{3/2}$ by an effective population temperature T^*

$$T^*=-\Delta E[\ln(N^*g_0/N_0g^*)]^{-1}, \quad (8)$$

where $\Delta E=2.1$ eV is the energy corresponding to the center of gravity of the doublet; g^* is its statistical weight; and N_0 and g_0 are the density and statistical weight of the sodium atoms in the ground state.

The analysis that follows is for conditions in the plasma that are realized for a relatively high content of inert gas in comparison with the total density of the alkali and the molecular admixture. In the range of electron temperatures under consideration $0.1\,\mathrm{eV}<T_e<1.5\,\mathrm{eV}$ it is possible to ignore processes of excitation of levels of the inert gas. Here and below, by T_e for the nonequilibrium EEDF we understand the quantity $T_e=(2/3)\langle\varepsilon\rangle$, where $\langle\varepsilon\rangle$ is the mean energy of the electrons. In the range of densities considered below, $[\mathrm{Na}]\sim[\mathrm{N_2}]$, the frequency of inelastic collisions with the sodium atoms significantly exceeds the frequency of electron excitation of the nitrogen molecules. In this regard, our calculations ignored the excitation of the $A^3\Sigma_u^+$ state and all higher-lying electron states of the nitrogen molecule. We ignored processes of excitation of all sodium levels lying above the $3P$ state from the $3S$ state. We did this because their excitation cross sections are at least an order of magnitude less than the cross section of the $3S-3P$ transition.

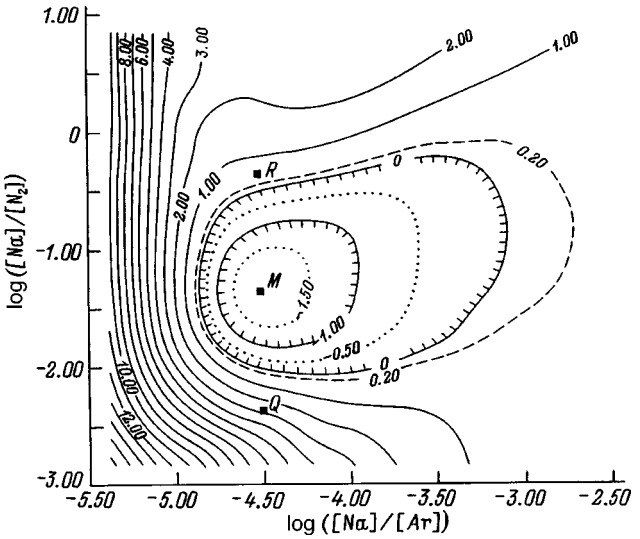

FIG. 2. Isolines of the electron mobility (in units of $10^{23}\,\mathrm{V}^{-1}\cdot\mathrm{cm}^{-1}\cdot\mathrm{s}^{-1}$) in a Na + Ar + N_2 mixture for zero degree of ionization plotted as a function of the partial ratios of nitrogen and argon in the mixture. The point M denotes the mobility minimum.

CALCULATED RESULTS

By way of an example, Fig. 2 plots the calculated electron mobility as a function of the partial ratios of the components of the mixture for $T^*=0.7$ eV. The calculation of the EEDF for these conditions neglected stepwise excitation from the $3P$ states of sodium and the Coulomb collisions, whose influence we discuss below. It can be seen that the electron mobility in a three-component mixture depends substantially on its composition. There are two regions of values of the electron mobility, in one of which its values are positive, and in the other, closed, region they are negative. As T^* decreases, the negative mobility existence region shrinks, and for $T^*<0.4$ eV the mobility takes only positive values.

Let us consider in more detail the conditions for the appearance of a negative electron mobility. As was shown in Ref. 30, there are two necessary (but insufficient) conditions for the appearance of the negative mobility effect. The presence of a negative derivative $\partial[\varepsilon^{3/2}/\nu_{el}(\varepsilon)]/\partial\varepsilon<0$. This criterion is met in the mixture under consideration only for argon in the energy region above the Ramsauer minimum. This inequality determines the fraction of argon atoms that

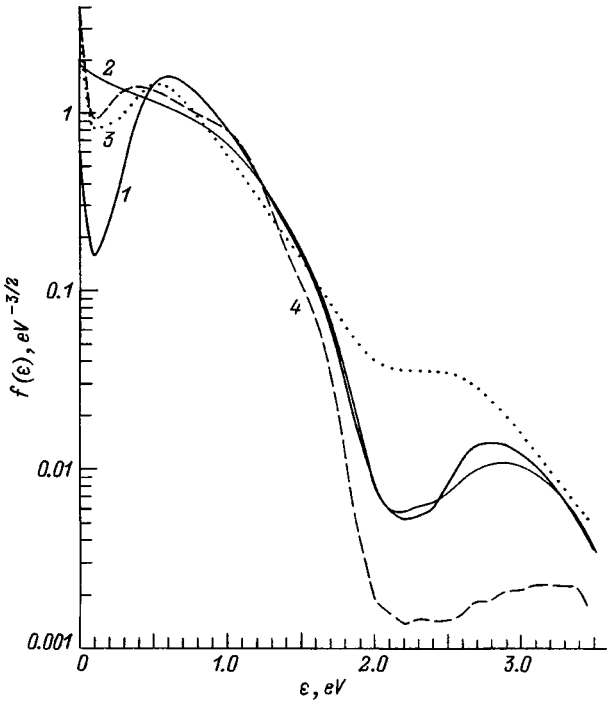

FIG. 3. Electron energy distribution function for different nitrogen concentrations and degrees of ionization. Ratios of the components of the mixture for curves *1–4* correspond to the points *M, M, R,* and *Q* (Fig. 2), respectively. The degree of ionization for curves *1, 3,* and *4* is equal to zero; for curve *2* it is equal to 10^{-9}.

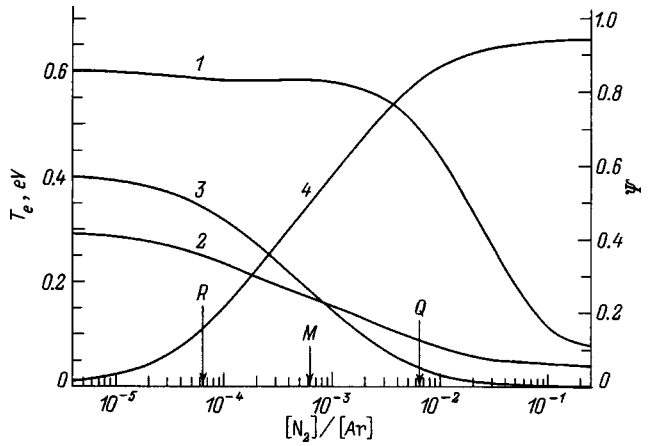

FIG. 4. Electron temperature and energy loss fraction (Ψ) for $T^*=0.7$ eV and different nitrogen concentrations in the mixture. The ratio [Na]/[Ar] corresponds to the minimum in the electron mobility (the point *M* in Fig. 2); *1* — T_e, *2–4* — energy loss fraction of the electrons for elastic and rotational collisions, inelastic collisions with Na, and vibrational excitation of levels of the $X^1\Sigma_g^+$ state of the N_2 molecule, respectively.

should exceed the total concentration of the two other components by three orders of magnitude. It is clear from Fig. 2 that in the region where the indicated condition is not fulfilled the electron mobility becomes positive. The maximum argon concentration is determined by the fact that the frequency of superelastic collisions with excited sodium atoms should exceed the frequency of energy losses in elastic collisions in the region of the Ramsauer minimum.[15] As can be seen from Fig. 2, when the argon concentration exceeds a certain value the mobility again becomes positive. This is due to a rapid relaxation of the electrons on elastic collisions with argon in the region of energies below the Ramsauer minimum. The second condition for existence of the effect of absolute negative conductivity is the presence of inversion in the EEDF $\partial f(\varepsilon)/\partial\varepsilon>0$.

Let us consider the role of nitrogen in the formation of inversion in the EEDF. Toward this end, we turn to Figs. 3 and 4, which plot results of a calculation of the EEDF and the electron temperature for various concentrations of the molecular admixture and fixed value of the ratio [Na]/[Ar] corresponding to the minimum point in the electron mobility. Figure 4 also plots the energy fractions going into excitation of molecular vibrations, inelastic losses in collisions with sodium atoms, and the total energy losses in elastic and rotational collisions.

For zero nitrogen concentration the main role in the formation of the EEDF is played by processes of excitation and de-excitation of the sodium atoms. As a result, a quasi-Maxwellian distribution is formed, with electron mobility near the equilibrium EEDF in argon for $T_e=T^*$ (Ref. 15). At the point *R* (see Figs. 2 and 3), where the relative nitrogen content is small, processes of vibrational excitation lead to the formation of a small inversion in the EEDF (Fig. 3, curve *3*). Here the mobility takes small (in comparison with the equilibrium EEDF, whose temperature is found from the equation of energy balance) but positive values.

With growth of the concentration of the molecular admixture (the point *M* in Figs. 2 and 3), the role of excitation of the vibrational levels grows. For nitrogen the excitation cross sections of the transitions $X^1\Sigma_g^+(v=0)\rightarrow X^1\Sigma_g^+(v)$ (v is the vibrational quantum number) fall rapidly with growth of v. Therefore, processes with energy losses of the same order of magnitude as the vibrational quantum of the N_2 molecule $E_1=0.29$ eV play the main role. It can be seen from Fig. 4 that growth of the efficiency of excitation of the vibrational levels is accompanied by a falloff of the inelastic losses in collisions with the sodium atoms. This is because the maximum of the vibrational excitation cross section corresponds to the near-threshold region of resonance transition energies of the sodium atom. In this range the excitation frequency of the molecular vibrations is comparable with the frequency of inelastic collisions with the sodium atoms. Since $E_1\ll\Delta E$, the energy balance is reached at almost the same mean energy of the electrons (Fig. 4). For the given ratios of the components of the mixture, a depletion of the EEDF in the region $\varepsilon_1<\varepsilon<\varepsilon_2$ is observed (Fig. 3, curve *1*). A decrease in the efficiency of inelastic collisions with sodium atoms in the ground state takes place since electrons are demoted to the subthreshold region of the $3S-3P$ cross section $\varepsilon<\Delta E$ as a result of excitation of molecular vibrations.

By leaving the efficient vibrational excitation zone $\varepsilon<\varepsilon_1$ the electrons relax preferably on elastic collisions with argon atoms. In the energy region corresponding to the Ramsauer minimum, relaxation into the low-energy range is retarded. Not succeeding in falling into the thermal energy region, the electrons undergo superelastic collisions with excited sodium atoms and return to the high-energy region. As

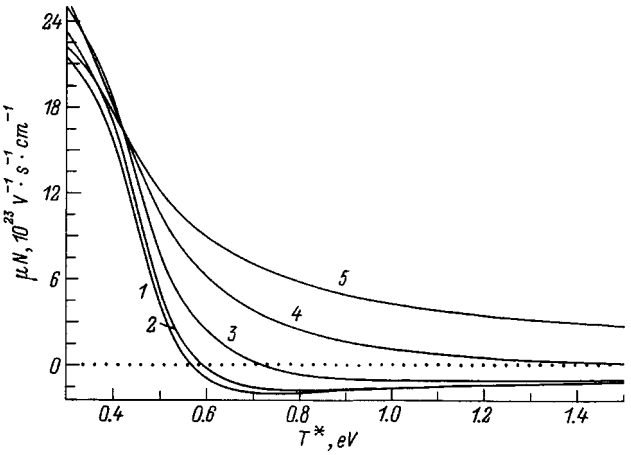

FIG. 5. Dependence of the electron mobility in a Na + Ar + N_2 mixture for different temperatures of excitation and degrees of ionization. The partial ratios of the components of the mixture correspond to the mobility minimum (the point *M* in Fig. 2). The degree of ionization for curves *1–5* is 0, 10^{-11}, 10^{-10}, 10^{-9}, 10^{-8}, respectively.

a result, a minimum is formed in the EEDF. Since an inverted distribution is formed in the region of the Ramsauer minimum, it makes the defining contribution to the integral (7) and the mobility becomes negative.

Further increase of the nitrogen density with all other conditions kept equal (the point *Q* in Figs. 2 and 4) increases the relaxation rate of the electrons on vibrational, rotational, and elastic collisions with nitrogen molecules. Superelastic collisions do not succeed in returning the electrons to the high-energy region, and a large part of the fast electron flux relaxes into the thermal energy region. The number of electrons in the low-energy region grows, and the inversion in the EEDF in the region of the Ramsauer minimum disappears (curve *4* in Fig. 3). In this case the mean energy of the electrons decreases and the mobility becomes positive.

The results of our calculations show that there is a composition of a three-component mixture at which the negative electron mobility reaches its maximum absolute value. This value and the composition of the mixture at which it is reached depend on the concentration of excited sodium atoms in the $3P$ state. However, the simultaneous necessity of a high concentration of excited atoms and small degree of ionization of the medium are contradictory requirements. Therefore, as our next step let us consider the problem of critical populations of the excited atoms and permissible electron concentrations at which the negative mobility effect can exist for the ratio of components corresponding to the point *M* in Fig. 2.

Taking Coulomb collisions into account decreases the width of the inversion interval in the EEDF since electron–electron collisions tend to build the equilibrium form of the energy distribution. In the final tally, an increase in the degree of ionization leads to destruction of the inversion in the EEDF (curve *2* in Fig. 3). Figure 5 plots the results of a calculation of the electron mobility as a function of the concentration of resonantly excited sodium atoms for different degrees of ionization of the medium. It can be seen that in a three-component mixture for degrees of ionization of the medium less than 10^{-9} the negative electron mobility effect is possible. The critical population of the excited atoms for degrees of ionization less than 10^{-10} corresponds to excitation temperatures from 0.55 to 0.7 eV. This corresponds to conditions in which less than 10% of the sodium atoms should be found in excited states. In comparison with a two-component mixture of Na and Ar (Ref. 15) the region of existence of negative mobility in a three-component mixture is expanded by two orders of magnitude in the degree of ionization and by one order of magnitude in the population of the excited sodium atoms.

In the above analysis we utilized the two-level model of the sodium atom and the assumption that $T_v = T_a$ for the equilibrium distribution of the molecules over the vibrational levels. This corresponds to conditions that can be realized for resonant optical excitation of the $3S-3P$ transition of sodium in the indicated mixture. However, under real experimental conditions various plasma-chemical processes can take place that alter the population of the N_2 vibrational levels and excited states of the sodium atoms. Here let us restrict the discussion to parametric calculations reflecting the influence of the indicated factors. Growth of the vibrational temperature leads to a decrease of inversion in the EEDF. There exists a critical T_v at which the mobility cannot take negative values. For the ratio of components of the mixture corresponding to the minimum value of the negative mobility in Fig. 2 for $T^* = 0.7$ eV the critical value of $T_v \sim 0.4$ eV.

Analysis shows that processes of stepwise population from the $3P$ state can play a substantial role in the formation of the EEDF for high $T^* \sim \Delta E$. The nearest levels $4S$ and $3D$, having the largest excitation cross sections, have the greatest influence. Reliable experimental data on these cross sections are still lacking. The available theoretical calculations disagree by an order of magnitude in the near-threshold energy region, which is the region having the maximum effect on the formation of the EEDF. In our calculations we used Ochkur's calculated cross sections.[31] In contrast to the cross sections given in the review article by Krishnan and Stumpf,[32] Ochkur's calculated cross sections allow for the exchange interaction as well as the dipole interaction, which leads to a growth of the corresponding cross sections in the near-threshold region by 40–60%. Calculation of the EEDF taking into account processes of stepwise excitation for conditions corresponding to the point *M* in Fig. 2 shows that for zero populations of the $4S$ and $3D$ levels the mobility does not take negative values. This is because stepwise excitation increases the electron flux into the thermal energy region and decreases the degree of inversion in the EEDF. Calculation shows that the EEDF is sensitive to the populations of the $4S$ and $3D$ levels. The existence range of the negative mobility is bounded from above by $T^* < \Delta E$. A detailed analysis of the influence of these processes could be had within the framework of a self-consistent model of a plasma, which goes beyond the scope of this work.

To sum up, we have analyzed the influence of nitrogen on the form of the EEDF in the three-component mixture N_2 + Ar + Na. We have shown that a nitrogen admixture increases the degree of inversion in the distribution function in the region of energies corresponding to the Ramsauer

minimum in the cross section of elastic collisions of the electrons with the argon atoms. We have found that the energy range of possible existence of negative electron mobility in the mixture N_2 + Ar + Na is extended into the region of smaller populations of the excited sodium atoms and higher degrees of ionization of the medium in comparison with an Ar + Na mixture.

We would like to express our deep gratitude to V. I. Ochkur for calculating the cross sections of stepwise excitation in sodium and to L. D. Tsendin for discussion of the results of this work.

[1] W. L. Highan, Phys. Rev. A **2**, 1989 (1970).
[2] R. I. Lyagushchenko and M. B. Tendler, Fiz. Plazmy **1**, 836 (1975) [Sov. J. Plasma Phys. **1**, 458 (1975)].
[3] N. L. Aleksandrov, A. M. Konchakov, and É. E. Son, Fiz. Plazmy **4**, 169 (1978) [Sov. J. Plasma Phys. **4**, 98 (1978)].
[4] A. Kh. Mnatsakanyan and G. V. Naĭdis, Fiz. Plazmy **2**, 152 (1976) [Sov. J. Plasma Phys. **2**, 84 (1976)].
[5] N. L. Aleksandrov, A. M. Konchakov, and É. E. Son, Fiz. Plazmy **4**, 1182 (1978) [Sov. J. Plasma Phys. **4**, 663 (1978)].
[6] C. Gorse, M. Cacciatore, and M. Capitelli *et al.*, Chem. Phys. **119**, 63 (1988).
[7] N. L. Aleksandrov and I. V. Kochetov, J. Phys. D **26**, 387 (1993).
[8] N. L. Aleksandrov, A. V. Dem'yanov, and I. V. Kochetov *et al.*, Fiz. Plazmy **23**, 658 (1997).
[9] A. G. Ponomarenko, V. N. Tishchenko, and V. A. Shveĭgert, Teplofiz. Vys. Temp. **25**, 787 (1987).
[10] F. Paniccia, C. Gorse, J. Bretagne, and M. Capitelli, J. Appl. Phys. **59**, 4004 (1986).
[11] N. A. Gorbunov, N. B. Kolokolov, and A. A. Kudryavtsev, Zh. Tekh. Fiz. **61**(6), 52 (1991) [Sov. Phys. Tech. Phys. **36**, 616 (1991)].
[12] N. A. Dyatko, I. V. Kochetov, and A. P. Napartovich, J. Phys. D **26**, 418 (1993).
[13] G. Colonna, C. Gorse, and M. Capitelli *et al.*, Chem. Phys. Lett. **213**, 5 (1993).
[14] H. Amemiya, S. Ono, and S. Teii, J. Appl. Phys. **56**, 4312 (1987).
[15] N. A. Gorbunov, F. E. Latyshev, and A. S. Mel'nikov, Fiz. Plazmy **24**, 950 (1998).
[16] N. A. Gorbunov, Ph. E. Latyshev, and A. S. Melnikov *et al.*, *23rd ICPIG*, Toulouse, 1997, Vol. 1, p. 80.
[17] N. A. Gorbunov, A. S. Melnikov, and I. A. Movtchan, *28th EGAS*, Berlin, 1997, p. 561.
[18] N. A. Gorbunov, F. E. Latyshev, and A. S. Mel'nikov *et al.*, FPPT-2, Minsk, 1997, Vol. 1, p. 9.
[19] T. Stacewicz and J. Krasinski, Opt. Commun. **39**, 35 (1981).
[20] L. D. Tsendin, Plasma Sources Sci. Technol. **4**, 200 (1995).
[21] N. A. Gorbunov, K. O. Iminov, and A. A. Kudryavtsev, Zh. Tekh. Fiz. **58**, 2301 (1988) [Sov. Phys. Tech. Phys. **33**, 1403 (1988)].
[22] A. I. Korotkov, A. A. Kudryavtsev, and N. A. Khromov, Zh. Tekh. Fiz. **66**(10), 92 (1996) [Tech. Phys. **41**, 1020 (1996)].
[23] Yu. P. Raĭzer, *Physics of Gas Discharge* [in Russian], Nauka, Moscow, 1987.
[24] L. G. H. Huxley and R. W. Crompton, *The Diffusion and Drift of Electrons in Gases* (Wiley, New York, 1974).
[25] I. P. Zapesochnyĭ, E. N. Postoĭ, and I. S. Aleksakhin, Zh. Éksp. Teor. Fiz. **68**, 1724 (1975) [Sov. Phys. JETP **41**, 865 (1975).
[26] A. V. Eletskiĭ, L. A. Palkina, and B. M. Smirnov, *Transport Phenomena in a Weakly Ionized Plasma* [in Russian], Atomizdat, Moscow, 1975.
[27] Sh. F. Araslanov, Dep. VINITI No. 2187-B87 [in Russian], All-Union Institute of Scientific and Technical Information, Moscow (1987), 68 pp.
[28] K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure*, Vol. 4, *Constants of Diatomic Molecules* (Van Nostrand, New York, 1979).
[29] M. G. Zhabitskiĭ and V. P. Silakov, MIFI Preprint No. 15-88 [in Russian], Moscow Engineering Physics Institute, Moscow (1988), 24 pp.
[30] N. A. Dyatko, I. V. Kochetov, and A. P. Napartovich, Pis'ma Zh. Tekh. Fiz. **13**, 1457 (1987) [Tech. Phys. Lett. **13**, 610 (1987)].
[31] V. I. Ochkur, private communication.
[32] U. Krishnan and B. Stumpf, At. Data Nucl. Data Tables **51**, 151 (1992).

Translated by Paul F. Schippnick

Excitation of bands of the first negative system of the N_2^+ ion in collisions of N^+ and O^+ ions with N_2 molecules

M. R. Gochitashvili and B. I. Kikiani

Tbilisi State University, 380028 Tbilisi, Russia
(Submitted August 2, 1996; resubmitted January 12, 1998)
Zh. Tekh. Fiz. **69**, 20–25 (April 1999)

Absolute values of the excitation cross sections of the (0,0) bands [for $O^+(^4S)$, $O^+(^2P)-N_2$ pairs] and the (0,0), (0,1), (1,2), and (2,3) bands [for $N^+(^3P)-N_2$ pairs] of the first negative system of the N_2^+ ion have been measured in collisions with nitrogen molecules of nitrogen and oxygen ions in the ground state and in a metastable state in the interval of ion energies $1-10$ keV. The process of excitation of the (0,0) band of the first negative system of the N_2^+ ion by oxygen ions in the metastable ^2P state is of a quasi-resonant character. The presence in the beam of ions in metastable states was monitored by measuring the excitation efficiency of the (0,0) band λ3914 Å of the N_2^+ ion in different operating regimes of the high-frequency ion source. For N^+ ions in the ^3P ground state, as the collision frequency is decreased the relative vibrational population of the $v'=1$ and $v'=2$ levels of the $B^2\Sigma_u^+$ state of the N_2^+ ion is observed to deviate strongly from the value calculated in the Franck–Condon model. © *1999 American Institute of Physics.* [S1063-7842(99)00404-3]

1. Excitation processes in ion–molecule collisions in the energy range from several hundred to several thousand electron volts are important for an understanding of phenomena taking place in the upper layers of the Earth's atmosphere,[1,2] in the interaction of particles of the solar wind with the atmospheric gases of the planets,[3] in a gas-discharge plasma, etc. One of the main sources of information about these phenomena is luminescence in the visible range of the spectrum induced in collisions.[4–6] In this regard, of first importance is information about processes of excitation of nitrogen atoms, molecules, and ions in collisions of electrons, protons, and ions of atomic nitrogen and oxygen with nitrogen molecules. In particular, data on cross sections of excitation of the bands of the first negative system (NS) of the N_2^+ ion (the transition $B^2\Sigma_u^+-X^2\Sigma_g^+$) in these collisions are needed.[7–10] This is because the luminescence of the indicated bands in the polar aurora is the most intense and is easily identified.[1] In addition, as was noted in Refs. 1 and 11, cascade transitions from higher-lying levels to vibrational levels of the $B^2\Sigma_u^+$ state of the N_2^+ ion are not observed and, consequently, such studies can give information about the population of these levels at the very instant of the collision event.

Results of a study of ion–molecule collisions in the indicated energy range are also of significant theoretical interest. At present there is no commonly accepted opinion of the mechanism of electron–vibrational excitation in these collisions. In particular, the question of the causes and nature of the deviation of the relative populations of the vibrational levels from the calculated values following from the Franck–Condon (FC) model has received insufficient study.[12]

In measurements of the cross sections of excitation of emission of molecules by N^+ and N^+ ion beams it is important to know in which states these ions are found. The reason for this, as was shown, for example, in Ref. 13, is that at ion energies $E\sim 1$ keV the cross sections of excitation of the bands of the first negative system of the N_2^+ ion by N^+ and O^+ ions in their collisions with N_2 molecules can differ by several orders of magnitude, depending on the state of the impinging ion.

The present paper reports results of measurements of cross sections of the excitation of the bands (0,0) λ3914 Å, (0,1) λ4278 Å, (1,2) λ4236 Å, and (2,3) λ4200 Å by the N^+ ion in the ^3P ground state and of the (0,0) band by the O^+ ion in the ^4S ground state and the metastable ^2P state in their collisions with N_2 molecules in the energy range $E=0.3-10$ keV.

2. The measurements were performed on a mass-spectrometer setup by means of optical spectroscopy. The experimental setup and the technique for calibrating the system for recording the optical radiation are described in Ref. 14: a primary particle beam from an rf source, accelerated to the required energy and focused, was mass-analyzed by a magnetic mass-analyzer, slit-collimated, and directed into the collision chamber filled with the investigated target gas. The measurements were carried out under single-collision conditions. The radiation emitted upon excitation of the molecules was observed at an angle of 90° with respect to the direction of the primary beam. A spectral analysis of the radiation was performed with the help of an MDR-2 monochromator.

The emission was recorded by an FÉU-79 photomultiplier in the integrating regime. With the aim of reducing the dark current of the photomultiplier, the photocathode was cooled by liquid-nitrogen vapors.

One peculiarity of the rf source used is that, depending on its operating regime, a certain fraction of the N^+ and O^+ ions are found in metastable states [the states $N^+(^1D,^1S,^5S)$ and $O^+(^2D,^2P)$]. Therefore, in those cases where it was

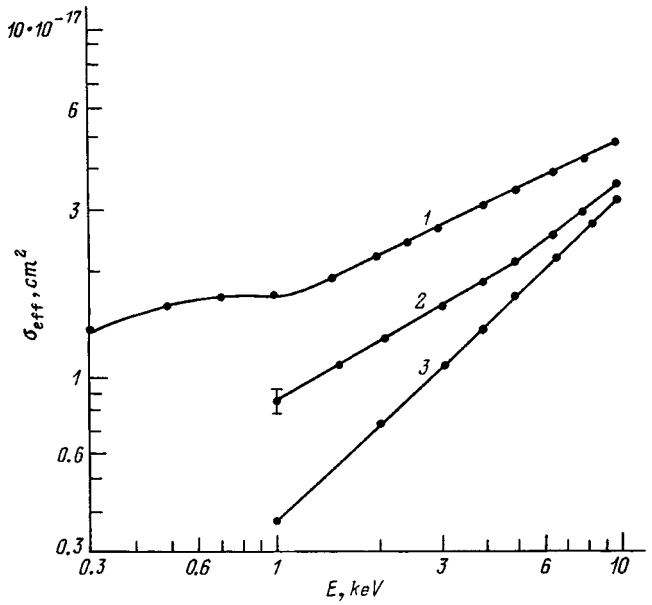

FIG. 1. Energy dependence of the efficiency of excitation of the band (0, 0) λ3914 Å of the first negative system of the N_2^+ ion for the pair $O^+ - N_2$ for different pressures of the working gas in the high-frequency source: *1* — $P = 2.4 \times 10^{-2}$, *2* — 8×10^{-2}, *3* — 2×10^{-1} Torr.

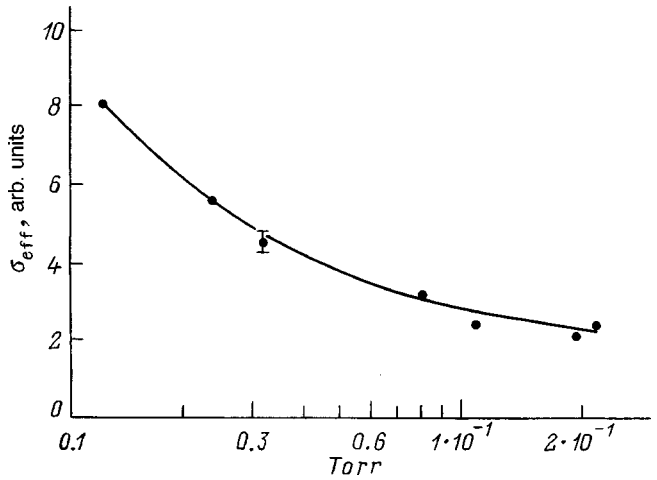

FIG. 2. Dependence of the efficiency of excitation of the (0,0) λ3914 Å band of the N_2^+ ion in $O^+ - N_2$ collisions on the working gas pressure in a high-frequency source ($E = 2.5$ keV).

needed we monitored the composition of the beam and determined the contribution to the efficiency of excitation of the bands of the first negative system of the N_2^+ ion by ions in metastable states.

3. Inelastic processes with N^+ and O^+ ions present in the primary particle beam in metastable states have been studied by a number of authors.[13,15,16] In particular, the authors of Ref. 13 measured the absolute cross sections of excitation of the (0,0) λ3914 Å band of the N_2^+ ion in collisions of N^+, O^+, and C^+ ions in the ground and metastable states with N_2 molecules at ion energies $E = 1$ keV.

From these measurements it follows that excitation of the indicated band by oxygen ions in the ^4S ground state proceeds with significantly lower efficiency than excitation by ions in the metastable ^2D and ^2P states (the ratio of the corresponding cross sections is 1:9:300). Consequently, even a small admixture in the primary particle beam of oxygen ions in metastable states can introduce a substantial error in the measured cross section of excitation of this band by ions in the ground state.

In the present study, the presence of ions in metastable states was monitored by measuring the efficiency of excitation of the (0,0) λ3914 Å bands of the N_2^+ ion for different operating regimes of the rf source. It follows from the measurements that in the case of $O^+ - N_2$ collisions the efficiency of excitation of the indicated band decreases significantly when the pressure of the working gas in the rf source is increased, and also when the voltage on the rf generator is decreased. This can be explained by a decrease in these cases of the fraction of O^+ ions in metastable states in the primary particle beam.

Figure 1 plots the energy dependence, obtained in the present work, of the efficiency of excitation of the (0,0) λ3914 Å band of the N_2^+ ion in collisions of the oxygen ion with nitrogen molecules at different pressures of the working gas in the 4f source: $P_1 = 2.4 \times 10^{-2}$ Torr, $P_2 = 8 \times 10^{-2}$ Torr, and $P_3 = 2 \times 10^{-1}$ Torr. By efficiency we mean the sum of terms, each of which is the product of the cross section of excitation of the given band by ions in the corresponding state by the relative fraction of these ions in the primary particle beam. As can be seen, these curves differ not only in their shape, but also in the overall magnitude of the excitation efficiency. For a fixed energy, as the pressure in the rf source is increased the excitation efficiency decreases.

Control experiments carried out in the present work show that at high enough pressures the fraction of ions in metastable states in the primary particle beam reaches a minimum. Results of these experiments are plotted in Fig. 2 in the form of a curve of the relative value of the efficiency of excitation of the indicated band plotted as a function of the oxygen pressure in the rf source at ion energies $E = 2.5$ keV. It can be seen that at pressures $P \geqslant 10^{-1}$ Torr the corresponding curve reaches saturation.

Comparison of our data with the data of Ref. 13 shows that for $E = 1$ keV and the oxygen pressure in the rf source $P_1 = 2.4 \times 10^{-2}$ Torr the excitation efficiency we measured ($\sigma_1 = 17 \times 10^{-18}$ cm^2, curve *1* in Fig. 1) is significantly greater than the values of the excitation cross section obtained in Ref. 13 for oxygen ions in the ^4S ground state ($\sigma = 0.73 \times 10^{-18}$ cm^2) and the metastable ^2D state ($\sigma = 6.6 \times 10^{-18}$ cm^2) significantly less than for ions in the metastable ^2P state ($\sigma = 220 \times 10^{-18}$ cm^2). Consequently, it can be surmised that the value of the excitation efficiency we obtained is determined to a significant extent by oxygen ions in the metastable ^2P state (for $E = 1$ keV).

In the present work the presence in the beam of oxygen ions in the metastable ^2D state was checked in a special control experiment. Toward this end, we examined excitation of the Meinel bands (3,0) λ6874 Å, (4,1) λ7037 Å, (5,2) λ7240 Å (the transition $A^2\Pi_u - X^2\Sigma_g^+$) of the N_2^+ ion in $O^+ - N_2$ collisions. It can be surmised that $O^+(^2D)$ ions will

excite this band with a large efficiency due to the quasiresonant character of the inelastic process

$$O^+(^2D)+N_2(X^1\Sigma_g^+)\rightarrow O(^3P)+N_2^+(A^2\Pi_u)$$

(energy defect $\Delta E\approx 0.06$ eV). The experiments showed that in our case the Meinel bands are hardly excited at all, which is probably an indication of the absence of oxygen ions in the metastable ^2D state in the primary particle beam. The small fraction of these ions in the primary particle beam in comparison with $O^+(^2P)$ ions is due to a peculiarity of the mechanism of formation of these ions during dissociative ionization in collisions of electrons with oxygen molecules in the rf ion source. An analysis of the energy terms of the oxygen molecule and the oxygen molecular ion shows that the most efficient mechanism of formation of dissociation products in various states is decay of highly excited states of the O_2^+ ion in the region of internuclear distances correspondin g to Franck–Condon transitions. Such states include $C^4\Sigma_u^-$, $III^2\Pi_u$, $B^2\Sigma_g^-$, $b^4\Sigma_g^-$, and $A^2\Pi_u$. In the decay of the $C^4\Sigma_u^-$ state of the O_2^+ ion, ions of atomic oxygen in the ^4S ground state and the metastable ^2P state are formed in the relative amounts ~95% and 5%, respectively.[17] Decay of the $III^2\Pi_u$ state of the O_2^+ ion leads to the formation of O^+ ions in the metastable ^2D and ^2P states in the relative amounts ~33 and 66%, respectively.[18] Decay of the $B^2\Sigma_g^-$, $b^4\Sigma_g^-$, and $A^2\Pi_u$ states leads to the formation of ions in the $O^+(^4S)$ ground state.[18] It follows from the foregoing discussion that oxygen ions in the ^4S ground state make up the main fraction while oxygen ions in the metastable ^2P state make up a small fraction of the oxygen ions in the primary particle beam and ions in the metastable ^2D state are essentially absent.

Using the data of Ref. 13, we can estimate from our results the fraction of oxygen ions in the ^4S ground state and in the metastable ^2P state in the primary particle beam for different values of the pressure of the working gas in the rf source. Simple calculations show that the fraction of $O^+(^2P)$ ions in the primary particle beam for a pressure of 2.4×10^{-2} Torr in the rf source is ~7.4%, for a pressure of 8×10^{-2} Torr it is ~3.4%, and for a pressure of 2×10^{-1} Torr it is 1.4%.

As was already mentioned, the $O^+(^2P)$ ions make a substantial contribution to the excitation efficiency as measured by us. Therefore, the decrease in the excitation efficiency brought about by an increase in the pressure inside the rf source is due entirely to a decrease in the fraction of these ions in the primary beam. The contribution of the metastable ions to the excitation efficiency also depends substantially on their energy — it decreases as their energy is increased.

As the energy of the oxygen ions is increased, the relative contribution of the ions found in the ^4S ground state to the excitation efficiency noticeably increases. As can be seen from Fig. 1, for the ion energy $E=10$ keV the change in the excitation efficiency with decrease of the pressure inside the rf source does not exceed ~30%.

After determining the fraction of metastable ions in the primary beam, it is easy to reconstruct the dependence of the cross section of excitation of the (0,0) λ3914 Å band by O^+ ions in the ^4S ground state and in the metastable ^2P state from the energy dependence of the excitation efficiency (Fig. 1) by solving some simple algebraic equations. The results of this reconstruction are shown in Fig. 3. It can be seen that the cross section of excitation of the given band by $O^+(^2P)$ ions grows slowly with increasing ion energy. Such an energy dependence of the cross section is not in accord with the quasiresonant character of excitation of the given band in the inelastic recharging process

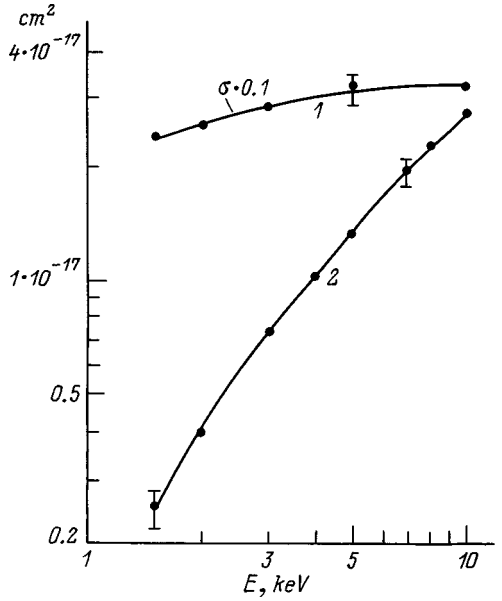

FIG. 3. Energy dependence of the excitation cross section of the (0, 0) λ3914 Å band for O^+-N_2 collisions: *1* — $O^+(^2P)$, *2* — $O^+(^4S)$.

$$O^+(^2P)+N_2(X^1\Sigma_g^+)\rightarrow O(^3P)+N_2^+(B^2\Sigma_u^+)(v'=0)$$

(energy defect $\Delta E\simeq 0.12$ eV). It is possible that this is the result of neglecting the contribution to the excitation efficiency from the O^+ ions found in the metastable ^2D state. As follows from our estimates, taking the contribution of the $O^+(^2D)$ ions to the excitation efficiency into account has only an insignificant effect (within the limits of experimental error) on the energy dependence of the cross section of excitation of the (0, 0) λ3914 Å band by O^+ ions in the ^4S ground state. This is due to the fact that at higher ion energies the contribution of the ions found in the ^4S ground state to the excitation efficiency becomes decisive.

An analogous procedure was carried out for the pair N^+-N_2. In this case the contribution of the nitrogen ions in metastable states to the excitation efficiency does not exceed ~30%. The energy dependence of the excitation efficiency is completely determined by the nitrogen ions in the ground state.

Final results of our measurements of the energy dependence of the absolute cross sections of excitation of the (0,0) λ3914 Å, (0,1) λ4278 Å, (1,2) λ4236 Å, and (2,3) λ4200 Å bands of the first negative system of the N_2^+ ion in collisions of N^+ in the ^3P ground state with N_2 molecules are plotted in Fig. 4.

An analysis of the obtained results shows that the influence of an admixture of ions in metastable states on the process of excitation of the molecular ion $N_2^+(B^2\Sigma_u^+)$ in N^+ O^+-N_2 collisions is substantial in the case of oxygen

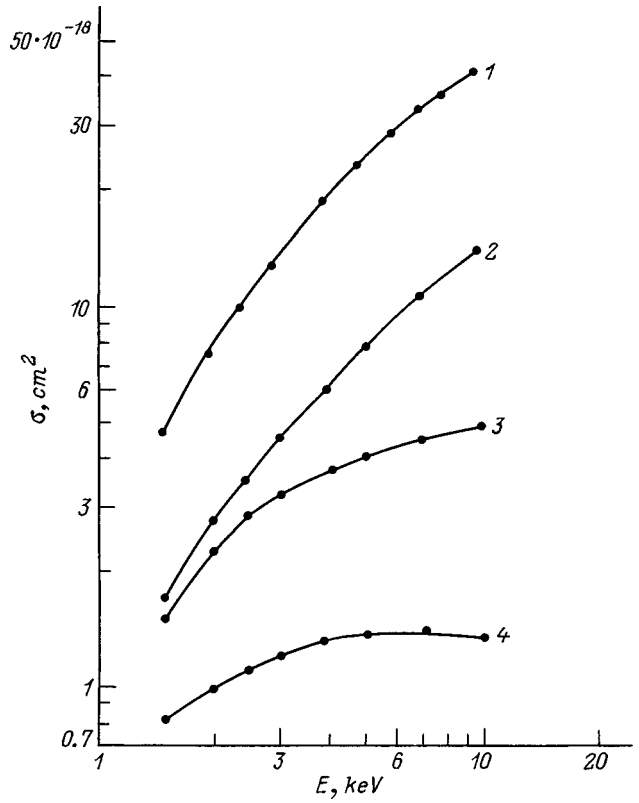

FIG. 4. Energy dependence of the excitation cross sections of the bands of the first negative system of the N_2^+ ion in $N^+(^3P)-N_2$ collisions: *1* — (0,0) λ3914 Å, *2* — (0,1) λ4278 Å, *3* — (1,2) λ4236 Å, *4* — (2,3) λ4200 Å.

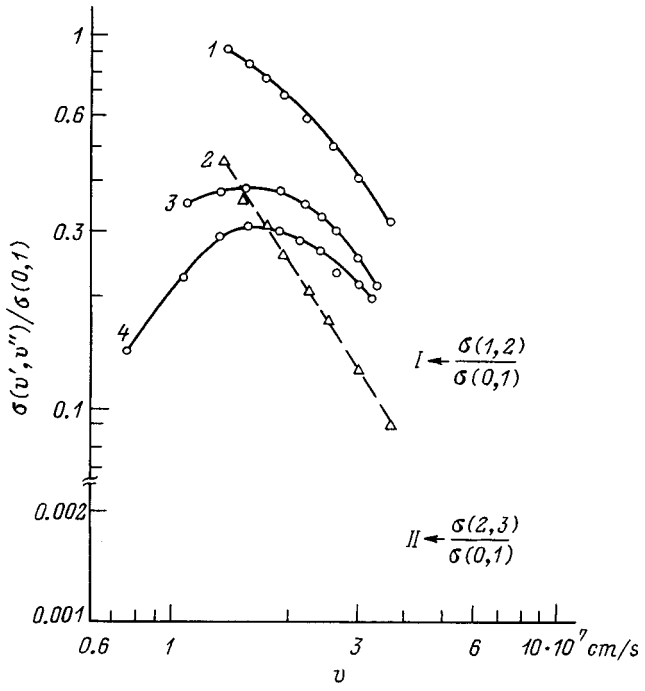

FIG. 5. Dependence of the ratio $\sigma(v',v'')/\sigma(0,1)$ of excitation cross sections of the bands of the first negative system of the N_2^+ ion on the velocity of the ions: *1* — $\sigma(1,2)/\sigma(0,1)$; *2* — $\sigma(2,3)/\sigma(0,1)$ for the pair N^+-N_2, the N^+ ions in the ^3P ground state; *3* — $\sigma(1,2)/\sigma(0,1)$ for the pair O^+-N_2, in the primary particle beam the $O^+(^2P)$ ion component is ~1.4%; *4* — $\sigma(1,2)/\sigma(0,1)$ for the pair O^+-N_2, in the primary particle beam the $O^+(^2P)$ ion component is ~7%.

ions in the low-energy end of the investigated energy interval and is small over the entire energy interval in the case of nitrogen ions.

Errors in the measurement of the absolute cross sections at the low-energy end of the investigated energy interval amount to ~30% and at the high-energy end, to ~20%. The errors in the relative measurements do not exceed ~5% over the entire interval of investigated energies.

4. The deviation of the populations of the vibrational levels during excitation of the $B^2\Sigma_u^+$ state of the N_2^+ ion in collisions of various ions with N_2 molecules from their values calculated according to the Franck–Condon model was investigated in earlier works.[9,19,20] In the present work this question is examined with allowance for the presence in the primary particle beam of ions in metastable states.

Our data for the particle pairs N^+, O^+-N_2 are plotted in Fig. 5 in the form of the dependence of the ratio of the excitation cross sections of the (1,2) and (2,3) bands to the excitation cross section of the (0,1) band on the collision velocity. For comparison, Fig. 5 also plots the values of these ratios calculated in the Franck–Condon model: $\sigma(1,2)/\sigma(0,1)=1.35\times10^{-1}$ (I) and $\sigma(2,3)/\sigma(0,1)=1.7\times10^{-3}$ (II).[12] These calculations used the Franck–Condon factors of the unperturbed states of the nitrogen molecule. Our results confirm the earlier observed deviation of the population of the excited vibrational levels from the value calculated in the Franck–Condon model: as the collision velocity is lowered, the relative population of the vibrational levels $v'=1$ and 2 increases in comparison with the calculated value.

Various models of electron-vibrational excitation in collisions of ions with N_2 molecules were discussed in Refs. 20–28. In the range of low and intermediate energies (from several hundred to several thousand electron volts) the formation of excited molecular states is due to a nonadiabatic transition between the potential energy surfaces of the initial and final states of the three-particle system of colliding particles in the region of minimal separation of the surfaces. The deviation of the relative populations of the high-lying vibrational levels from the value calculated in the Franck–Condon model is possibly due to a violation of the assumptions made in the calculation: the electron transitions occur instantaneously for a fixed value of the internuclear distance in the molecule when the relative population of the vibrational levels of the final electron state is determined by the Franck–Condon factors of the unperturbed molecule and is consequently independent of the collision velocity and the nature of the impinging ion. However, as was shown in Refs. 20–26, in the region of minimal separation of the potential energy surfaces of the three-particle system of colliding particles vibrational excitation of the molecules takes place, leading to a redistribution of the populations of the vibrational levels of the final state. Of course, it goes without saying that in these cases the use of the Franck–Condon factors for the unperturbed molecule requires special care. In particular, as was pointed out in these papers, the degree of deviation of the relative populations from their values calcu-

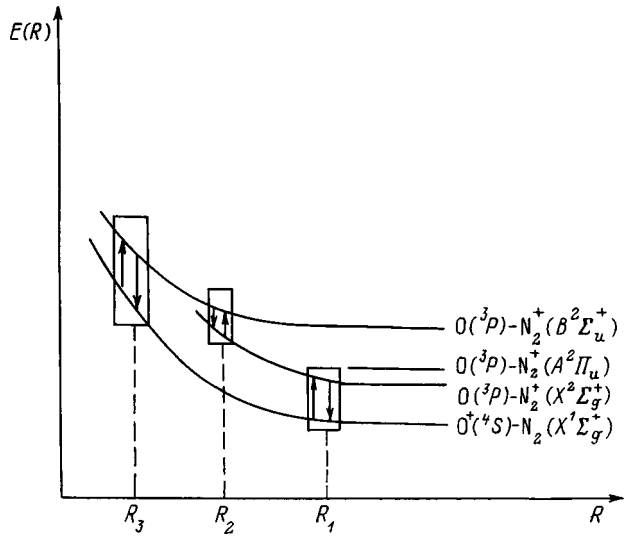

FIG. 6. Schematic diagram of the potential curves of the system $(ON_2)^+$: R is the distance between the primary particle and the center of mass of the molecular target; R_1, R_2, and R_3 are the R values of closest approach of the curves of the initial and intermediate, intermediate and final, and initial and final states, respectively.

lated in the Franck–Condon model depends on the collision velocity, the specific form of the potential energy surfaces, and the quantum number of the excited vibrational level. To be specific, the degree of deviation grows with decrease of the collision velocity and increase of the vibrational quantum number v'. A large deviation of the relative populations from their calculated values is observed in both cases. For example, under conditions of close collisions, when the impact parameter is of the order of the internuclear distance in the molecule, momentum transfer of significant magnitude is possible to an individual nucleus of the molecule. In this case the transition to the excited molecular state originates from a vibrationally excited initial molecular state. The same result can be obtained if the system transitions to the final state directly through an intermediate state[20,23] or if the system in the initial state receives a vibrational excitation due to a strong exchange interaction (causing a stretching of the molecule) with some other near-lying state.[24] In this case the degree of vibrational excitation and, consequently, the deviation of the relative population from its value calculated in the Franck–Condon model depend substantially on the residence time of the system in the intermediate or vibrationally excited initial state.[23,24]

In our case, the main reason for the deviation of the relative population of the vibrational levels in the N^+-N_2 collision from its value calculated in the Franck–Condon model is the factor of efficient momentum transfer under conditions of close collisions (the cross sections we measured were small) between identical nuclei — the impinging N^+ ion and one of the two N atoms in the N_2 molecule.

Of particular interest, in our opinion, are the results for the O^+-N_2 pair. It can be seen from Fig. 5 that in this case a significant deviation of the relative vibrational distribution from its value calculated in the Franck–Condon model is observed only for a comparatively small admixture ($\sim$1.4%) of O^+ ions in the metastable ^2P state (Fig. 5, curve *3*). With increase of the fraction of ions in metastable states in the primary particle beam the population of high-lying vibrational levels decreases, especially for small collision velocities. For example, for the collision velocity $v\sim 1.1\times 10^7$ cm/s, as the fraction of $O^+(^2P)$ ions in the primary particle beam is increased from $\sim$1.4% to $\sim$7% the ratio $\sigma(1,2)/\sigma(0,1)$ decreases by roughly a factor of two. The point here is, as was already mentioned, that the process of excitation of the $B^2\Sigma_u^+$ state of the N_2^+ ion in collisions of $O^+(^2P)$ ions with N_2 molecules has a quasiresonant character and consequently occurs for large distances R between the ion and the center of mass of the molecule. Here it is natural to assume that a weak perturbation of the N_2 molecule can be expected and momentum transfer to one of its atoms does not take place. Therefore, for $O^+(^2P)$ ions a distribution of the relative populations of the vibrational levels similar to the distribution calculated according to the Franck–Condon model is expected. As the velocity of the ions is decreased, as was already pointed out, the contribution of the metastable ions to the efficiency of excitation of the $B^2\Sigma_u^+$ state of the N_2^+ ion grows and the relative population of the vibrational levels should correspondingly approach the distribution calculated in the Franck–Condon model. Indeed, for the ion velocity $v=0.76\times 10^7$ cm/s, when the number of oxygen ions in the metastable ^2P state in the primary particle beam is comparatively large, $\sim$7% (Fig. 5, curve *4*), the ratio $\sigma(1,2)/\sigma(0,1)$ is 0.14, which is close to the value calculated in the Franck–Condon model. For this velocity the ratio $\sigma(1,2)/\sigma(0,1)$ is determined entirely by O^+ ions in the metastable ^2P state. It follows from the above discussion that a deviation of the relative vibrational population in O^+-N_2 collisions from its calculated value is only observed for O^+ ions in the ^4S ground state. This result can be explained by an analysis of the diagram of potential energy curves of the system $(ON_2)^+$ constructed in the quasidiatomic approximation.[20,29]

For the case of $O^+(^4S)-N_2(X^1\Sigma_g^+)$ collisions, the redistribution of vibrational levels (relative to their values calculated in the Franck–Condon model) in the excited molecular state $B^2\Sigma_u^+$ of the N_2^+ ion is probably explained by the influence of the $O(^3P)-N_2^+(X^2\Sigma_g^+)$ intermediate state. In this case, an electronic transition occurs via the given intermediate state or the vibrationally excited initial molecular state $N_2(X^1\Sigma_g^+)$. The latter can be vibrationally excited as a result of an exchange interaction of the initial state with the intermediate state.

We may sum up by enumerating the main results of the present work. In collisions of N^+ and O^+ ions in their ground states we have measured the absolute excitation cross sections of the (0,0) band (for the pair O^+-N_2) and the (0,0), (0,1), (1,2), and (2,3) bands (for the pair N^+-N_2) of the first negative system of the N_2^+ ion in the interval of ion energies 1–10 keV. For N^+ ions in the ^3P ground state, as the collision velocity is decreased a strong deviation of the relative vibrational population of the final state from its value calculated in the Franck–Condon model is observed while for O^+ ions in the metastable ^2P state the measured and calculated values of the vibrational population distribution are similar.

In conclusion, we would like to express our gratitude to V. M. Lavrov for his interest in this work.

[1] J. W. Chamberlain, *Physics of the Aurora and Airglow* (American Geophysical Union, Washington, 1995).
[2] M. A. Geller, H. Volland, H. Mayr and I. Harris, *The Polar Upper Atmosphere* [trans. of Pt. I of *Exploration of the Polar Upper Atmosphere*, edited by C. S. Deehr and J. A. Holtet, D. Reidel Publ. Co., Hingham, Mass.; Kluwer, Boston (1981); Mir, Moscow (1983), 456 pp.].
[3] V. A. Krasnopol'skiĭ, *Physics of the Glow of Planetary and Cometary Atmospheres* [in Russian], Nauka, Moscow, 1987, 304 pp.
[4] D. R. Bates, Proc. R. Soc. London, Ser. A **A**, 217 (1949).
[5] R. L. Gattinger and A. V. Jones, Can. J. Phys. **52** (23), 2343 (1974).
[6] O. C. Cartwright, J. Geophys. Res. **83A**, 517 (1978).
[7] P. N. Stanton and R. M. St. John, J. Opt. Soc. Am. **59**, 252 (1969).
[8] F. J. De Heer and J. F. M. Aarts, Physica **48**, 620 (1970).
[9] J. H. Birely, Phys. Rev. A **10**, 550 (1974).
[10] V. Degan, J. Quant. Spectrosc. Radiat. Transf. **18**, 1113 (1977).
[11] F. R. Gilmore, J. Quant. Spectrosc. Radiat. Transf. **5**, 369 (1965).
[12] R. W. Nicholls, J. Quant. Spectrosc. Radiat. Transf. **2**, 433 (1962).
[13] Ch. Ottinger and J. Simonis, Chem. Phys. **28**, 97 (1978).
[14] M. R. Gochitashvili, R. V. Kvizhinadze, N. R. Dzhaliashvili, and B. I. Kikiani, Zh. Tekh. Fiz. **63**(10), 35 (1993) [Tech. Phys. **38**, 851 (1993).
[15] B. M. Hughes and T. O. Turner, J. Chem. Phys. **55**, 3419 (1977).
[16] M. Vujovic, M. Matic, B. Cobic, and P. Hvelplund, J. Phys. B **10**, 3699 (1977).
[17] T. Akahari, Y. Morioka, M. Watanabe *et al.*, J. Phys. B **18**, 2219 (1985).
[18] F. B. Yousif, B. G. Lindsey, F. R. Simpson, and C. J. Latimer, J. Phys. B **20**, 5079 (1987).
[19] J. H. Moore Jr. and J. P. Doering, Phys. Rev. A **177**, 218 (1969).
[20] V. M. Lavrov, M. R. Gochitashvili, V. A. Ankudinov, and B. I. Kokiani, Zh. Tekh. Fiz. **50**, 660 (1980) [Sov. Phys. Tech. Phys. **25**, 400 (1980)].
[21] J. D. Kelley and M. Wolfsberg, J. Chem. Phys. **53**, 2967 (1970).
[22] J. D. Kelley, G. H. Bearman, H. H. Harris, and J. J. Leventhal, Chem. Phys. Lett. **50**, 295 (1977).
[23] E. A. Gislason, A. W. Kleyn, and J. Los, Chem. Phys. **59**, 91 (1981).
[24] E. A. Gislason and E. M. Goldfield, Phys. Rev. A **25**, 2002 (1982).
[25] U. C. Klomp, M. R. Spalburg, and J. Los, Chem. Phys. **83**, 33 (1984).
[26] D. Dhuicq and V. Sidis, J. Phys. B **20**, 5089 (1987).
[27] M. Lipeles, J. Chem. Phys. **51**, 1252 (1969).
[28] S. M. Fernandez, F. J. Erikson, A. V. Bray, and E. Pollack, Phys. Rev. A **12**, 1252 (1975).
[29] O. Yenen, D. H. Jaecks, and P. J. Martin, Phys. Rev. A Rev. A **35**, 1517 (1985).

Translated by Paul F. Schippnick

Current–voltage characteristics of underneutralized Knudsen thermionic converters with electron emission from the collector

V. I. Sitnov, A. Ya. Énder, and E. V. Yakovlev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted July 28, 1997)
Zh. Tekh. Fiz. **69**, 26–34 (April 1999)

Calculations of the potential distribution[4] are used to study the current–voltage (I–V) characteristics of unneutralized Knudsen thermionic converters with unlimited electron emission from the collector. It is shown that the optimum work function of the collector satisfies the condition of unlimited emission. The effect of a number of characteristic parameters of a converter on its I–V characteristics are studied: the degree of neutralization, the ratio of the emitter and collector temperatures, and the work functions of the emitter and collector.
 [S1063-7842(99)00504-8]

INTRODUCTION

The efficiency of thermionic converters of thermal into electrical energy is highest in the Knudsen regime, when the mean free path of the electrons exceeds the interelectrode gap. In this regime the emitter is simultaneously a source of electrons and ions (the electrons are produced by thermionic emission and the ions by surface ionization), and the thermionic converter is a Knudsen diode with surface ionization. The advantage of the Knudsen regime compared to the arc regime for a thermionic converter is the absence of the energy losses in and voltage drop across the working gap which are necessary in the arc in order to generate ions. There is an absolute need for more efficient thermionic converters in modern power sources for both space- and ground-based applications.

In practice, the Knudsen regime can be realized either by using a binary Cs–Ba filling[1] or by building a converter with a small gap.[2]

In thermionic converters with Cs–Ba, the barium vapor is used for independent control of the emitter work function and, because of the high adsorption energy of barium, it is possible to obtain an optimum emitter work function, while remaining in the Knudsen regime for interelectrode gaps d of about 1 mm. On the other hand, when the gap is reduced it is possible to raise the cesium vapor pressure while remaining in the Knudsen regime. For $d \sim 20$ μm an optimum emitter work function at a temperature of 1800–2000 K is ensured by adsorption of cesium on the emitter surface. In this case, the Knudsen regime is realized without adding barium vapor. For a fixed temperature of the collector, its optimum work function in the Knudsen regime is such that the electron emission from its surface is comparable to the electron emission form the emitter surface. We note also that a high collector emissivity is realized in thermionic converters with a Cs–Ba filling because the collector temperature must be kept higher than the barium reservoir temperature ($\sim$1000 K) corresponding to an optimum barium vapor pressure. Therefore, it is important to take the reverse current from the collector into account when studying the distributions of the potential from the current–voltage (I–V) characteristics of thermionic converters.

The first studies of the effect of electron emission from the collector were made by McIntyre.[3] However, because of the large number of parameters and the complexity of the calculations, no systematic data were obtained, and the calculations were done only for a few examples and are actually only useful for a rough estimate of the effect of emission from the collector. A more detailed study of the distributions of the potential in thermionic converters with unneutralized emitting collectors has been made.[4]

This paper is devoted to a study of the I–V characteristics, including a detailed analysis of the singular points on these characteristics, and also to a study of the effect of various parameters on the I–V characteristics.

1. One of the important parameters determining the operation of a thermionic converter in the Knudsen regime is the degree of neutralization

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

Here $n_i^+(0)$ and $n_e^+(0)$ are the densities of the ions and electrons escaping from the emitter immediately at its surface. In the unneutralized regime $\gamma < 1$. Besides γ there are a number of other parameters which must be taken into account in studies of thermionic converters: the ratio $\Theta = T_E/T_C$ of the emitter and collector temperatures, and the work functions χ_E and χ_C of the emitter and collector measured in units of kT_e.

In Ref. 4 the calculations were done assuming unlimited emission from the collector, i.e., it was assumed that the emission from the collector is so high that the flux of electrons entering the plasma is limited only by the potential barrier near the collector and not by its emissivity. It was found that, as opposed to thermionic converters with a non-emitting collector, in this regime there are no oscillating potential distributions. The evolution of the potential distributions as the external voltage u is varied is shown in Fig. 1. As the voltage is increased, there is a transition from a potential

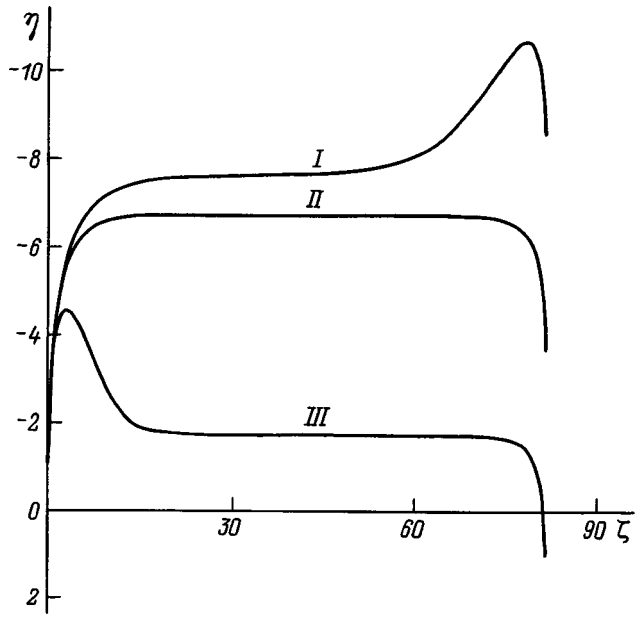

FIG. 1. The possible types of potential distribution (ζ is the distance from the emitter measured in Debye radii).

distribution with a virtual collector (anode) (curve *I* of Fig. 1) to one with a virtual emitter (cathode) (curve *III*) through an intermediate type (curve *II* of Fig. 1). At the peak of the virtual collector (or virtual emitter) the potential reaches a minimum value in the gap. Here and in the following, we shall use dimensionless values of the potentials η, measured in units of kT_E/e (where k is Boltzmann's constant and e is the charge of the electron). We denote the minimum potential in the interelectrode gap by η_m and the plasma potential by η_p. We shall denote the parameters corresponding to the transition from potential distributions with a virtual collector to those with a virtual emitter by the index α. The point α has been analyzed in detail before.[4]

In calculating the I–V characteristics we use the method described in Ref. 4. First η_p and η_m are found at the characteristic points on the potential distribution. Two conditions serve for finding these potentials: quasineutrality in the plasma and zero total charge in the layer between η_m and η_p. In order to determine the densities of the corresponding groups of particles (ions from the emitter, electrons from the emitter, and electrons form the collector) at an arbitrary point, we have introduced the function F_s. The quasineutrality condition is written in the form

$$(n_i-n_e)|_{\eta=\eta_p}=F(\eta_p,\eta_m,\Theta)=\gamma F_i(\eta_p) - F_{eE}(\eta_p,\eta_m)-\beta F_{eC}(\eta_p,\eta_m,\Theta)=0. \quad (2)$$

Here n_i and n_e are the densities of the ions and electrons, respectively; $\beta=n_e^-(\eta_m)/n_e^+(0)$; $n_e^-(\eta_m)$ is the electron density leaving the collector at the peak of the virtual emitter (or virtual collector). The contributions to the charge are determined through the functions G_s, which for an arbitrary gap (A,B) are are defined as the integral of F_s over the given gap interval. The condition that the total charge in the layer $[\eta_p,\eta_m]$ equal zero has the form

$$G(\eta_p,\eta_m,\Theta)=\gamma G_i(\eta_p,\eta_m)-G_{eE}(\eta_p,\eta_m) - \beta G_{eC}(\eta_p,\eta_m,\Theta)=0. \quad (3)$$

The parameter β depends on η_m as[4]

$$\beta=\Theta^{-3/2}\exp[(\eta_m-u)\Theta-\chi_E(\Theta-1)]. \quad (4)$$

The forms of the functions F_s and G_s are given in Ref. 4.

By specifying the external voltage u, it is possible to determine η_p and η_m from the solution of Eqs. (2)–(4) and, from the magnitude of the minimum potential, to calculate the transferred current. Detailed studies of the I–V characteristics will be made in the following: Fig. 2 shows a family of I–V characteristics for different collector work functions

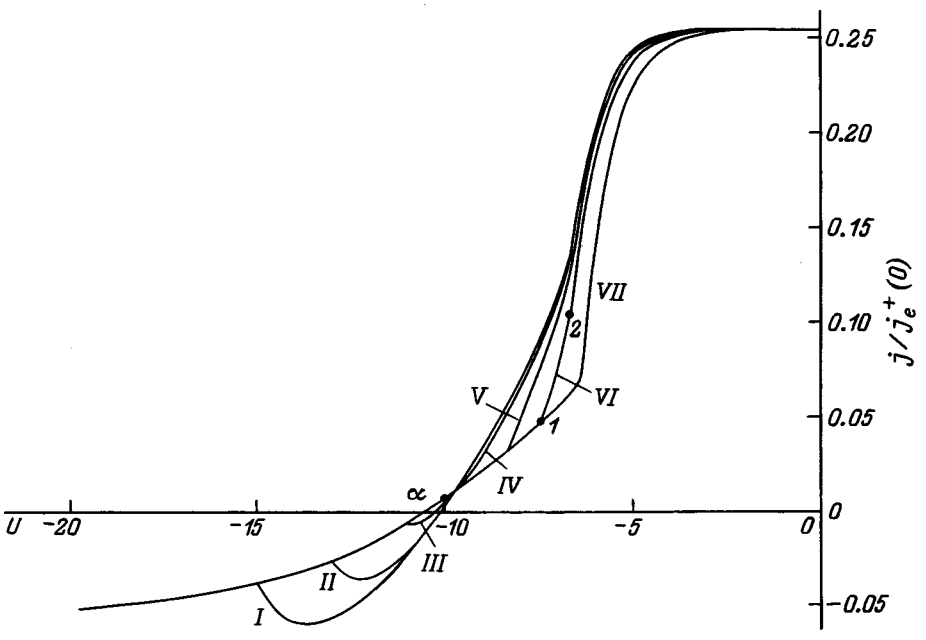

FIG. 2. I–V characteristics for different χ_C: *I* — 0.0, *II* — 2.0, *III* — 4.0, *IV* — 5.3, *V* — 6.7, *VI* — 7.6, *VII* — 8.6; γ=0.1, χ_E=15, Θ=2.

χ_C. On the I–V characteristic corresponding to $\chi_C=7.6$ (curve *VI*), three characteristic points are indicated: the point α, as well as points corresponding to the passage through zero of the collector potential η_C (point *1*) and the plasma potential (point *2*).

2. As a rule, the points *1* and *2* lie to the right of the point α on the I–V characteristic, so we shall concentrate the analysis specifically on the region $u>u_\alpha$, which corresponds to a potential distribution with a virtual cathode. It has been shown[5] that for the quasineutrality of the plasma and the collector potential jump to match, the following condition must be satisfied:

$$\frac{d}{d\eta}(n_i-n_e)|_{\eta=\eta_p+0}<0. \tag{5}$$

$\eta=\eta_p+0$ has been chosen because for a potential distribution with a virtual emitter, the transition from a quasineutral plateau in the anode region must be accompanied by an increase in the potential, i.e., we must have $\eta_C \geqslant \eta_p$. We shall refer to the condition (5) as the differential criterion.

In examining solutions including the reflection of ions from the collector sheath, we shall use yet another, more rigid criterion for the correctness of the solution. Let us consider an arbitrary point with potential η and the function $G(\eta)$ given by

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

Here and in the following, for convenience we shall omit the arguments η_m and Θ in the functions F and G. The function $G(\eta)$ is related to the electric field strength at the point with potential η, $E(\eta)$, by the equation[4]

$$G(\eta)=-E^2(\eta)/2. \tag{7}$$

It follows from Eq. (7) that the function G must be less than or equal to zero, i.e.

$$G(\eta)\leqslant 0. \tag{8}$$

We shall refer to Eq. (8) as the integral criterion. Equation (8) has to be used along with the differential criterion for the following reasons: criterion (5) is local and makes it possible to verify the correctness of the solution only in the nearest neighborhood of the boundary point of the quasineutrality region. Since $n_i=n_e$ in the quasineutrality region, when condition (5) is satisfied the anode layer near the plateau is enriched in electrons, i.e., $n_i<n_e$. However, when the collector potential is positive, the total density (n_i-n_e) can change sign in the anode region and, although Eq. (5) is satisfied, the total charge in the layer between the plasma and some point in the anode region (the function G) turn out to be positive.

The I–V characteristics were calculated assuming unlimited emission from the collector. Evidently, for high collector work functions χ_C this condition fails. For each point on the I–V characteristics we can find a limiting value χ_C^∞ corresponding to unlimited emission. This value is found from the condition $\eta_C=\eta_p$, since if $\chi_C>\chi_C^\infty$, then $\eta_C<\eta_p$ and the unlimited emission condition is violated. We shall do the calculations assuming that $\eta_C=\eta_p$. Calculations[4] show that at the point α the plasma potential is negative. Therefore, even in the neighborhood of that point, for $\chi_C=\chi_C^\infty$ the inequality $\eta_C<0$ holds, i.e., ions are not reflected from the potential jump near the collector. However, if χ_C is small, it may turn out to be positive and in this case it is necessary to take reflection of the ions into account. We denote the value of the work function of collector corresponding to the condition $\eta_C=0$ by χ_C^r; if $\chi_C<\chi_C^r$, then ions are reflected from the potential jump at the collector. Both χ_C^∞ and χ_C^r depend on u.

Therefore, each point on the I–V characteristic corresponds to a range of values of the collector work function, $[\chi_C^r,\chi_C^\infty]$, such that if χ_C lies within this interval, then the condition of unlimited emission is satisfied and the ions are not reflected. In this case it is not necessary to change the expression for F_i in Eq. (2) and, therefore, the corresponding formula for G_i. This, in turn, means that if the calculations have been done assuming $\eta_C=\eta_p(\chi_C=\chi_C^\infty)$, then η_p and η_m are the same for all χ_C within this interval. In this case the I–V characteristic is independent of χ_C, i.e., it is universal.

Note that similar considerations can be applied to the region to the left of the point α. Since in this case the collector is in contact not with the quasineutrality region but with a virtual collector, the work function χ_C^∞ corresponds to having $\eta_C=\eta_m$; as before, the value of χ_C^r corresponds to having $\eta_C=0$. The parameters χ_E, χ_C, u, and η_C are related by the simple equation[4]

$$\chi_E-\chi_C+u=\eta_C. \tag{9}$$

For a given value of u, χ_C^r and χ_C^∞ are found by substituting $\eta_C=0$ and $\eta_C=\eta_p$ in Eq. (9) for a virtual emitter ($\eta_C=0$ and $\eta_C=\eta_m$ for a virtual collector); $\eta_p(u)$ and $\eta_m(u)$ are found by solving Eqs. (2)–(4). It is clear from Fig. 3, which shows χ_C^r and χ_C^∞ as functions of the voltage u, that for large negative voltages, the interval $[\chi_C^r,\chi_C^\infty]$ is quite wide. Then, the χ_C^r are small.

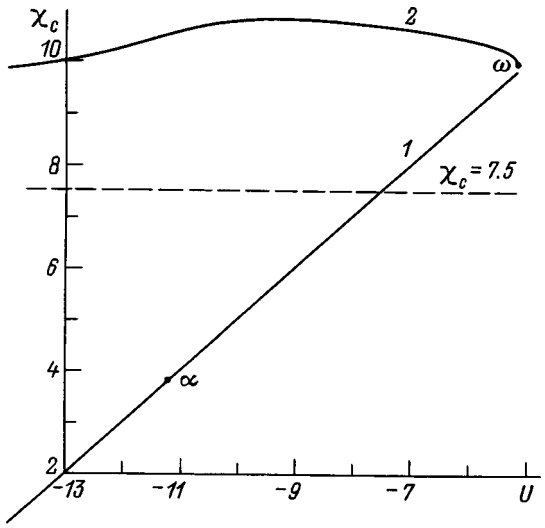

FIG. 3. χ_C^r (*1*) and χ_C^∞ (*2*) as functions of u for $\gamma=0.01$, $\chi_E=15$, and $\Theta=2$.

When constructing the universal I–V characteristic in the region of a virtual emitter, the work function of the collector was not assumed to be constant, but for each u the value of χ_C was chosen to be equal to χ_C^∞, in order to satisfy the condition $\eta_C=\eta_p$. At each point of the I–V characteristic a test was made to see if criterion (5) is satisfied. The point where condition (5) is violated is the boundary for existence of solutions of this type. We shall denote its parameters by the index ω. The state ω is determined from the condition that for $\eta_C=\eta_p$,

$$\frac{d}{d\eta}(n_i-n_e)\big|_{\eta=\eta_p}=0. \tag{10}$$

At this time the derivative changes sign, and for $u>u_\omega$ no solutions with a potential plateau exist. The parameters of the point $\omega-\eta_{m\omega}$, $\eta_{p\omega}$ and u_ω, can be found using the condition (10). Adding condition (10) to Eqs. (2) and (3), we obtain a system of three equations which can be written in the form

$$A_1\gamma+B_1\beta+C_1=0,\quad A_2\gamma+B_2\beta+C_2=0,$$
$$A_3\gamma+B_3\beta+C_3=0 \tag{11}$$

with the coefficients

$$A_1=\mathrm{exers}(-\eta_p),\quad B_1=\mathrm{exers}(\Theta\delta)-2\ \exp(\Theta\delta),$$
$$C_1=-\exp(\eta_m)\cdot\mathrm{exers}(\delta);$$
$$A_2=-\mathrm{exers}(-\eta_p)+\frac{1}{(-\pi\eta_p)^{1/2}},$$
$$B_2=\Theta\cdot\mathrm{exers}(\theta\delta)-2\ \exp(\Theta\delta)-\frac{1}{(\pi\Theta\delta)^{1/2}},$$
$$C_2=\exp(\eta_m)\cdot\left(\mathrm{exers}(\delta)-\frac{1}{(\pi\delta)^{1/2}}\right);$$
$$A_3=Ir(-\eta_p)-Ir(-\eta_m),$$
$$B_3=-\frac{1}{\Theta}(Ir(\theta\delta)+1-2\ \exp(\Theta\delta)),$$
$$C_3=\exp(\eta_m)\cdot(Ir(\delta)-1).$$

Here $\delta=\eta_p-\eta_m$, $\mathrm{exers}(x)=\exp(x)\cdot(-\mathrm{erf}(x^{1/2}))$,

$$\mathrm{erf}(x)=\frac{2}{\pi^{1/2}}\int_0^x\exp(-t^2)dt,\quad Ir(x)=\int\mathrm{exers}(x)dx.$$

In Eq. (11) the coefficients with subscript 1 correspond to the functions F_s, those with subscript 2, to derivatives of the F_s, and those with subscript 3, to the functions G_s. The solution of Eq. (11) gives the parameters of the point ω. Some calculated parameters of the point ω corresponding to $\chi_E=15$ are listed in Table I.

Calculations of the I–V characteristic in the virtual emitter region on the assumption that $\eta_C=\eta_p$ showed that η_p is negative along the entire segment of the I–V characteristic from the point α to the point ω for arbitrary γ. Thus in the virtual emitter region there is always a range of the collector work function, $[\chi_C^r,\chi_C^\infty]$, for which the I–V characteristic is universal.

TABLE I.

	$\Theta=2$			$\Theta=3$		
	u_ω	$\eta_{p\omega}$	$\eta_{m\omega}$	u_ω	$\eta_{p\omega}$	$\eta_{m\omega}$
$\gamma=0.01$	−5.15	−0.10	−4.82	−8.65	−0.05	−4.73
$\gamma=0.1$	−6.27	−0.12	−2.61	−9.39	−0.06	−2.51

The state ω corresponds to an extreme value of the collector work function, $\chi_{C\omega}^\infty$. Given our assumption that $\eta_C=\eta_p$, we can obtain an expression for $\chi_{C\omega}^\infty$ from Eq. (9): $\chi_{C\omega}^\infty=\chi_E-\eta_{p\omega}+u_\omega$.

If χ_C is much greater than $\chi_{C\omega}^\infty$, the condition for unlimited emission is violated for all u. When χ_C is slightly greater than $\chi_{C\omega}^\infty$, in the region (u_α,u_ω) there is a small range of voltages within which this condition is satisfied. Evidently, when the condition for unlimited emission is violated, oscillatory potential distributions develop. As χ_C increases, there is a transition to the developed oscillatory potential distributions typical of operation with a nonemitting collector.

Let us see to what extent the condition for unlimited emission is satisfied in the optimum Knudsen regime for a thermionic converter with Cs–Ba filling. This sort of thermionic converter has been optimized without self-consistent accounting for the effect of the electrons emitted from the collector on the potential distribution.[1] Let us consider a typical example for a high temperature thermionic converter: $T_E=2200$ K, $T_C=1100$ K ($T_E/T_C=2$), gap length $d=1$ mm, and optimum cesium vapor pressure $P_{\mathrm{Cs}}=2\times10^{-2}$ Torr. In this case, the optimum electrode work functions were $\chi_E^0=3.42$ eV (18.06 kT) and $\chi_C^0=1.85$ eV (9.77 kT). The ratio $\chi_E^0/\chi_C^0=1.85$, which is somewhat smaller than T_E/T_C. For the given χ_E^0 with $\gamma=0.2$, $\chi_{C\omega}^\infty$ equals 10.05, and with $\gamma=0.1$, $\chi_{C\omega}^\infty=10.38$. Therefore, if the collector work function is less than the optimum, then in the underneutralized regime the condition for unlimited emission is obviously satisfied. An analysis showed that this conclusion is valid for other temperatures of the emitter and collector, as well.

3. We now proceed to a study of the portion of the I–V characteristic where ion reflection is important. The I–V characteristics must be calculated for fixed values of the collector work function χ_C. In order for the condition of unlimited emission to be satisfied in the virtual emitter region, it is necessary that χ_C be less than $\chi_{C\omega}^\infty$. It is clear from Fig. 3 that for $u=u_\omega$ the value of χ_C^r, which we shall denote by $\chi_{C\omega}^r$, is very close to $\chi_{C\omega}^\infty$. If we choose a value of χ_C less than $\chi_{C\omega}^r$, then for some u the value of χ_C turns out to be equal to χ_C^r. As an example, the dashed line $\chi_C=7.5$ in Fig. 3 intersects the $\chi_C^r(u)$ curve at $u=-7.5$. At this time the collector potential η_C changes sign and ion reflection sets in.

I–V characteristics were calculated for a number of values of χ_C within the interval $\chi_{C\alpha}^r<\chi_C<\chi_{C\omega}^r$. $\chi_{C\alpha}^r$ is found from the condition that at the time when the plasma passes through state α, the potential of the collector surface passes through zero. For $\eta_C=0$ and $u=u_\alpha$, Eq. (9) yields $\chi_{C\alpha}^r$

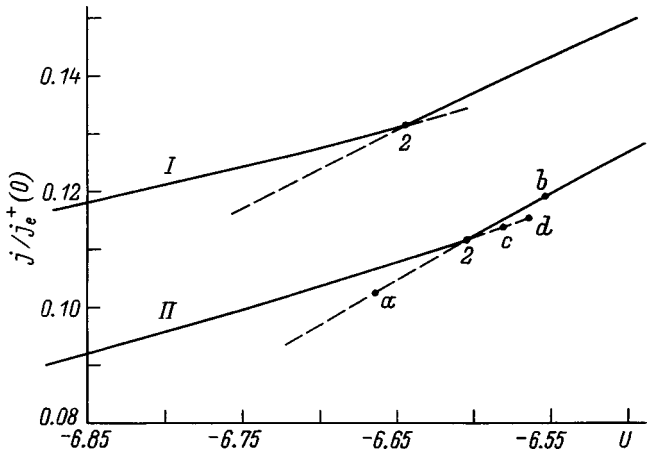

FIG. 4. An example of a segment of a I–V characteristic for $\gamma=0.1$, $\chi_E=15$, and $\Theta=2$ in the neighborhood of point *2* for two values of χ_C: *I*—5.5 and *II* — 7.5.

$=\chi_E+u_\alpha$. Including ion reflection makes the formula for the ion density depend on η_C:

$$F_i(\eta,\eta_C)=\begin{cases}2\,\mathrm{exers}(-\eta)-\exp(-\eta_C)\\ \quad\times\mathrm{exers}(\eta_C-\eta), \quad \eta<0,\\ 2\,\exp(-\eta)-\exp(-\eta_C)\\ \quad\times\mathrm{exers}(\eta_C-\eta), \quad \eta\geqslant 0.\end{cases} \tag{12}$$

The formula for G_i also changes. As before, the I–V characteristic was constructed by solving the system of Eqs. (2)–(4) using the appropriate formulas for F_i and G_i. At the time η_C passes through zero, the ion density increases rapidly and additional neutralization of the electron space charge occurs, so that there is a sharp rise in the current and a bend is seen in the I–V characteristic (point *1* of Fig. 2). Here the voltage at point *1*, u_1, is found from the condition

$$u_1=\chi_C-\chi_E, \tag{13}$$

which corresponds to the contact potential difference.

In Fig. 2 one can see the envelope and the parts of the I–V characteristic corresponding to the different values of χ_C which are characteristic of all I–V characteristics. The work function of the collector influences the I–V characteristic, despite the unlimited emission from the collector, because of ion reflection and the strong effect of the ion density on the potential distribution and on η_m.

4. The following feature of the I–V characteristic shows up in the neighborhood of the transition through zero of the plasma potential η_p. During the calculation each point of the I–V characteristic was checked to see if the differential criterion (5) was satisfied there, i.e., the sign of $d(n_i-n_e)/d\eta$ was determined. The calculations showed that for small negative values of η_p a situation arises in which $d(n_i-n_e)/d\eta$ changes sign, i.e., there is an approach to an extreme solution of the type which occurs at the point ω, with the difference that in this case, this solution takes into account the reflections of ions from a barrier at the collector. The corresponding point is an extremum in the lower branch of the I–V characteristic (Fig. 4) which corresponds to a negative plasma potential.

Besides solutions with a negative η_p, there are solutions with a virtual emitter for positive η_p (the upper branch of the I–V characteristic). An analysis showed that the extremum point on this branch is determined by a solution with $\eta_p=0$. At this point the differential criterion is violated when the outer boundary of the virtual emitter and the quasineutrality region are matched. In the neighborhood of the extrema in the upper and lower branches there is a range of u where two solutions of the system of Eqs. (2)–(4) can be found. These solutions are formal and it is still necessary to verify that the integral criterion (8) is satisfied for them.

The upper and lower branches always intersect. We shall denote the point of intersection as point *2*. The portion of a I–V characteristic in the neighborhood of point *2* is shown in Fig. 4. All the possible formal solutions are indicated in this figure. At point *2*, itself, for a given current (i.e., for a single value of η_m) and a single collector potential, there are two solutions with different values of η_p of which the smaller is negative and the larger, positive.

We now apply criterion (8) to verify the correctness of the formal solutions in the neighborhood of point *2*. It is evident that criterion (8) must be satisfied for all points in the anode region, as well as for the points in the outer portion of a virtual emitter. If at some point between η_m and η_C the function $G(\eta)>0$, then the corresponding potential distributions cannot exist. Detailed studies showed that criterion (8) is violated only in a small neighborhood of point *2*.

Let us consider the solutions corresponding to the upper branch of the I–V characteristic in the neighborhood of point *2*. In this case, criterion (8) is violated in the outer part of the virtual emitter. Figure 5b shows plots of $F(\eta)$ and $G(\eta)$ in the interval (η_m,η_p) corresponding to point *2* on the I–V characteristic for $\chi_C=7.5$. It is clear that, besides the point η_p, there is yet another point within this interval where F and G go simultaneously to zero. The potential at this point is negative and equal to the plasma potential for the solution corresponding to the lower branch at point *2*. For a small shift to the left along the upper branch from point *2* $G(\eta)$ is positive within some range of values of η lying within this interval; that is, the integral criterion is violated and the formal solution is not a solution to the problem. Functions $F(\eta)$ and $G(\eta)$ of this type, which correspond to the point *a* in Fig. 4, are shown in Fig. 5a. Calculations showed that over the entire segment of the I–V characteristic to the left of point *2*, the solutions do not satisfy criterion (8). To the right of point *2*, G is negative and the integral criterion is not violated, as illustrated in Fig. 5b, which corresponds to point *b* of Fig. 4.

In the upper branch at point *2* and to the left of it, the function $F(\eta)$ goes to zero three times within the interval (η_m,η_p) (Fig. 5a and 5b). The zeros of F correspond to extrema of G. The two outermost extrema are minima of G, while the middle extremum is a maximum. If we choose a point on the upper branch lying to the right of point *2*, then two variants are possible: either $F(\eta)$ has three zeroes within the interval, as before, and all the extrema of G are negative (this case occurs immediately adjacent to point *2*), or $F(\eta)$ goes to zero once and G has a negative minimum at the corresponding point (Fig. 5c) and is always increasing on

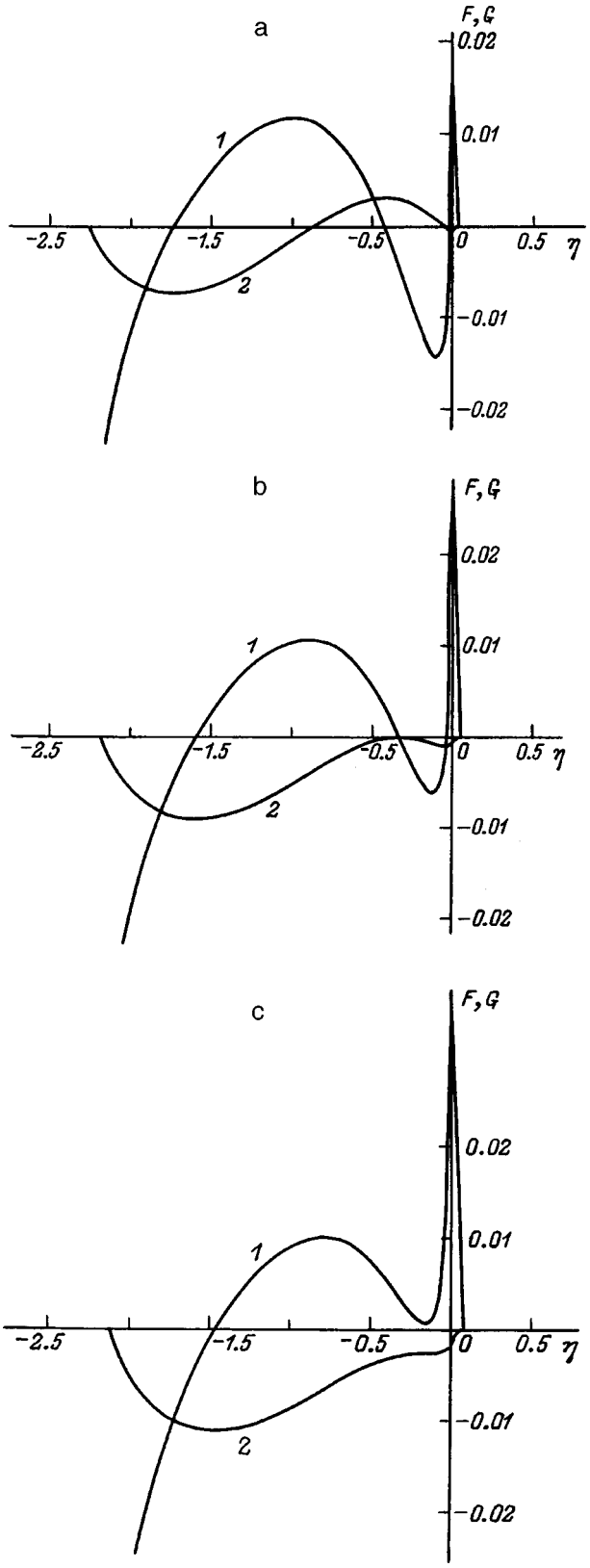

FIG. 5. F (*1*) and G (*2*) as functions of η for the upper branch when $\gamma=0.1$, $\chi_E=15$, $\Theta=2$, and $\chi_C=7.5$; $u=6.66$ (a), 6.60 (b), and 6.55 (c).

the segment from the minimum point to the point η_p. In both cases, at the right boundary of the interval (η_m, η_p) the function G goes to zero by definition. The boundary point on the upper branch, u^*, corresponds to the function G going to

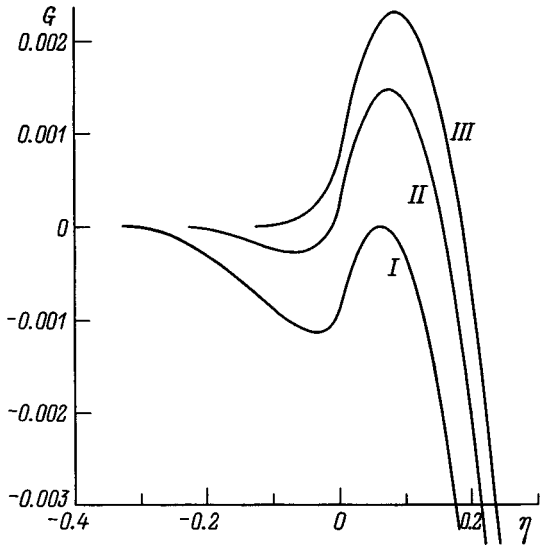

FIG. 6. Plots of $G(\eta)$ for the lower branch with $\gamma=0.1$, $\chi_E=15$, $\Theta=2$, and $\chi_C=7.5$; $u=-6.60$ (*I*), -6.58 (*II*), and -6.56 (*III*).

zero at the maximum, and this means that here the functions F and G go simultaneously to zero; i.e., this is a potential distributi on with a plateau whose potential lies within the interval (η_m, η_p). The calculations showed that the potential of this plateau is always negative. Thus, for a given η_m, there are two solutions with different plasma potentials and the point u^* corresponds to the intersection point of the upper and lower branches (point *2*).

A similar examination of the solutions on the lower branch showed that to the right of point *2*, $G(\eta)$ becomes positive in the neighborhood of the jump at the collector. Figure 6 shows plots of $G(\eta)$ for different points on the lower branch of the I–V characteristic for $\eta \geqslant \eta_p$. Curve *I* in Fig. 6 corresponds to point *2* and curves *II* and *III* correspond to points *c* and *d* in Fig. 4, which lie on the I–V characteristic to the right of point *2* with curve *III* belonging to the extremum point on the lower branch. It is clear from curves *II* and *III* that there is a region where the function G is positive. Our calculations show that to the left of point *2* on the lower branch the integral criterion is satisfied everywhere. Thus, the formal solutions of the system of Eqs. (2)–(4) can not be regarded as solutions to our problem everywhere, even when condition (5) is satisfied: on the lower branch the solutions lie to the left of point *2* and on the upper branch, to the right. The formal solutions of Eqs. (2)–(4) which do not satisfy the integral condition are indicated by the dashed lines in Fig. 4. The I–V characteristic is single valued and continuous; at point *2* there is only a bend in the I–V characteristic, while the plasma potential changes discontinuously at this point.

The curves $F(\eta)$ and $G(\eta)$ in Figs. 5 and 6 are for $\gamma=0.1$, $\chi_E=15$, $\Theta=2$, and $\chi_C=7.5$. An examination of these curves over a wide range of these parameters showed that they have a similar form.

The effect of χ_C on the I–V characteristic was studied over a wide range of variation in the degree of neutralization γ. As an example, Fig. 2 shows the calculations for $\gamma=0.1$ (curves *IV*–*VII*).

If $\chi_C < \chi^r_{C\alpha}$, then ion reflection and the departure of the I–V characteristic from the envelope set in to the left of point α and the transition from a virtual collector to a virtual emitter occurs on the branch with reflection. Figure 2 shows I–V characteristics for a number of values of χ_C smaller than $\chi^r_{C\alpha}$ (curves *I–III*). It is clear that the I–V characteristic to the right of the reflection point moves to some part below the envelope. This is because of a change in the relative contribution to the total current from the flux of collector electrons to the emitter compared to the flux of emitter electrons to the collector.

5. When emission takes place from the collector, the I–V characteristic depends on many parameters: γ, Θ, χ_E, χ_C, and T_E. As the above remarks imply, T_E enters only as a unit for measuring the potentials, while χ_C affects the I–V characteristic only because of ion reflections.

Let us consider the dependence of the I–V characteristic on the emitter work function. We assume that the calculations have been done and the I–V characteristic constructed for some fixed value of $\chi_E = \chi^f_E$. It has been shown[4] that the voltage at point α is given by

$$u_\alpha = (\eta_\alpha - \chi_E)\cdot(1 - 1/\Theta) - \ln\Theta/\Theta, \tag{14}$$

while at the point α the potential η_α is independent of the emitter work function χ_E and is determined only by γ and Θ. On going to a new value of χ_E, η_α remains unchanged, while the voltage at point α is given by

$$u_\alpha = u^f_\alpha = -(\chi_E - \chi^f_E)\cdot(1 - 1/\Theta). \tag{15}$$

Here u^f_α is the voltage at point α for $\chi_E = \chi^f_E$. It can be shown that the other points of the segment of the I–V characteristic where there is no ion reflection also shift to the same value. This simplifies the problem of calculating the I–V characteristic for other χ_E.

Suppose the I–V characteristic has been calculated for some value $\chi_C = \chi'_C$ of the collector work function. Then the onset of ion reflection corresponds to point *1*, at which the voltage is $u'_1 = \chi'_C - \chi^f_E$ [in accordance with Eq. (13)]. In order for point *1* to move the same distance as point α on going to the new value of χ_E (i.e., in order for the distance between these two points to remain unchanged), it is necessary to change the collector work function so that

$$\chi_C - \chi'_C = (\chi_E - \chi^f_E)/\Theta. \tag{16}$$

It can be shown that, in the ion reflection region, for these two I–V characteristics (one constructed for χ^f_E and χ'_C, the other for χ_E and χ_C) at points located the same distance from point α, the potential distribution and, therefore, the transferred currents are the same. Thus, if as χ_E varies we vary χ_C according to Eq. (16), then the I–V characteristic shifts as a whole along the u axis by an amount $u_\alpha - u^f_\alpha$ and all the calculations can be done for a fixed value of $\chi_E = \chi^f_E$. For the following calculations we have chosen $\chi^f_E = 15$.

In the calculations shown below the collector work function was chosen to be close to optimal from the condition $\chi_E/\chi_C = T_E/T_C$, [1] i.e., $\chi_C = \chi_E/\Theta$. It is interesting to note that in this case u_1 lies within the range (u_α, u_ω), and is always roughly in the middle of this interval.

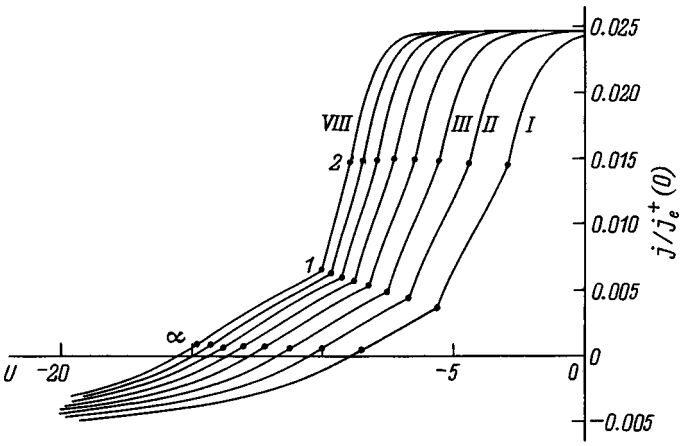

FIG. 7. I–V characteristics for different Θ: *I* — 1.6, *II* — 1.8, *III* — 2.0, *IV* — 2.2, *V* — 2.4, *VI* — 2.6, *VII* — 2.8, *VIII* — 3.0; γ=0.01, χ_E=15, and $\chi_C = \chi_E/\Theta$.

Figure 7 illustrates the effect of Θ on the I–V characteristic. Studies showed that for all γ, as Θ is reduced the characteristic shifts to the right and the current at point *2* is almost unchanged while the current at point *1* drops. As u is increased, the curve approaches saturation. The magnitude of the saturation current is independent of Θ and equals the corresponding quantity for a nonemitting collector. This is easy to understand, since when the collector voltage is increased there is a large jump in the potential and electrons from the collector cannot overcome this barrier and affect the potential distribution near the emitter.

When studying the effect of γ on the I–V characteristic it is convenient to take the current relative to the saturation current j_s. The corresponding I–V characteristics are shown in Fig. 8. It is clear that, on one hand, reducing γ causes an increase in the absolute magnitude of the open circuit voltage u_{xx}, while, on the other, for currents close to the saturation current, it causes the I–V characteristic to shift to the right. As a result, I–V characteristics corresponding to different γ and a single value of Θ intersect one another, with the intersection point always lying at currents between 0.2 and 0.3 and between the singular points *1* and *2*. The existence of such a universal point for I–V characteristics with different γ may be useful for the analysis of experimental I–V characteristics.

Studies showed that the I–V characteristics are very similar to one another in the reverse current region. If we superimpose the open circuit points for a fixed value of Θ and different γ, then the I–V characteristics in the reverse current region will be close to one another. Thus, one of our major conclusions is the following: in an underneutralized thermionic converter with unlimited electron emission from the collector, a change in the saturation current in the forward branch should be accompanied by a proportional change in the current in the reverse branch. This proportionality exists because the electronic charges in the virtual emitterregion (on the forward branch) and in the virtual collector region (on the reverse branch) are neutralized by the same ions. A comparison between the reverse current portions of experimental I–V characteristics from thermionic converters

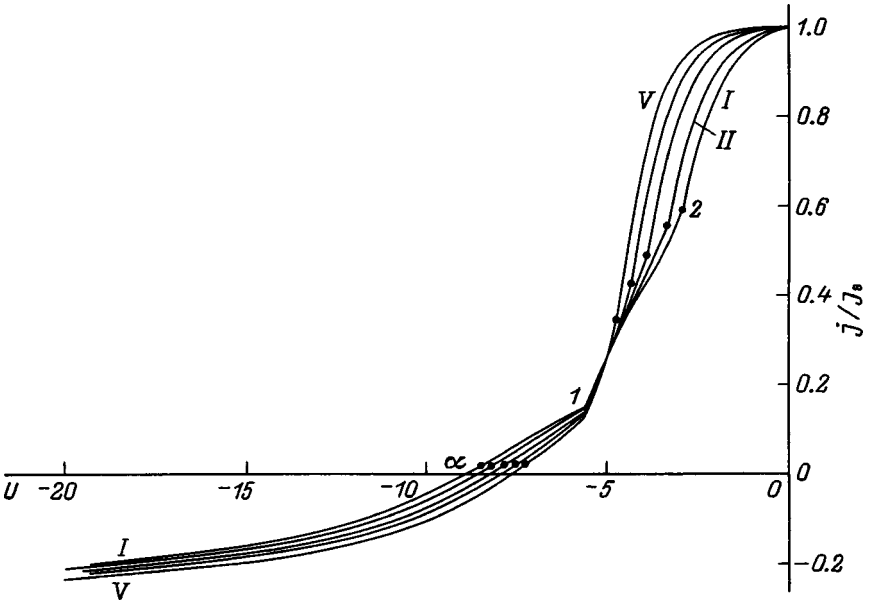

FIG. 8. I–V characteristics for different γ: *I* — 0.01, *II* — 0.02,*III* — 0.05, *IV* — 0.1, *V* — 0.2; $\chi_E = 15$, $\Theta = 1.6$, and $\chi_C = \chi_E/\Theta = 9.375$.

with a Cs–Ba filling and the theoretical curves confirms this conclusion.

CONCLUSION

We have performed systematic calculations of the I–V characteristics of thermionic converters in the underneutralized regime assuming unlimited electron emission from a hot collector. We have analyzed a number of singular points on the I–V characteristic associated both with the onset of ion reflection from the collector potential jump and with the plasma potential's passing through zero. It has been shown that the I–V characteristic is everywhere single valued and continuous. The effect of various parameters on the form of the I–V characteristic has been investigated.

The calculations were done subject to a number of assumptions. As usual, it was assumed that all the charged particles which reach one or another electrode are absorbed there. It was assumed that there is no trapping of charged particles in a potential well. This assumption has been discussed and justified elsewhere.[4]

It was, furthermore, assumed that the electrodes have a uniform work function. The electrodes actually have a spotty structure. This should lead to some difference between the experimental and theoretical I–V characteristics. All of the bends obtained near the singular points of the theoretical I–V characteristics are smoothed out. For example, ion reflection begins when the spot on the collector with the lowest work function has a potential equal to zero. As the collector potential is raised, spots with a higher work function will open up and the ion reflection process will become smooth. Nevertheless, a comparison of experimental I–V characteristics with the theoretical ones can be used to evaluate the major parameters of the thermionic converter operating regime and can then be used for quantitative diagnostics of hot-collector, Knudsen thermionic converters.

This work was supported by the Russian Fund for Fundamental Research (Project No. 97-02-18080).

[1] V. I. Babanin, V. I. Kuznetsov, A. S. Mustafaev *et al.*, Zh. Tekh. Fiz. **48**, 754 (1978) [Sov. Phys. Tech. Phys. **23**, 444 (1978)].
[2] R. Ya. Kucherov and Y. V. Nikolaev, in *Proceedings of the 30th IECEC* (1995), Paper No. CT 416, pp. 141–144.
[3] R. G. McIntyre, Proc. IEEE **51**, 760 (1963).
[4] V. I. Sitnov and A. Ya. Énder, Zh. Tekh. Fiz. **68**(4), 37 (1998) [Tech. Phys. **43**, 382 (1998)].
[5] A. Ya. Ender, Candidate's Dissertation [in Russian], Leningrad (1972), 242 pp.

Translated by D. H. McNeill

Self-consistent development and fractal structure of leader discharges along a water surface

V. P. Belosheev

S. I. Vavilov State Optical Institute All-Russia Science Center, 199034 St. Petersburg, Russia
(Submitted October 21, 1997; resubmitted March 10, 1998)
Zh. Tekh. Fiz. **69**, 35–40 (April 1999)

Ideas regarding the development of single- and multichannel leader discharges over water surfaces are generalized on the basis of experimental data.[1,2] The Ohmic conductivity of water is found to be manifested in the self consistency of their dynamics and fractal structure. The fractal dimensionality of a single-channel leader discharge is found to be 0.96 ± 0.05 and that of a multichannel discharge to be 1.85 ± 0.05. Mechanisms are proposed for the branching of leader discharge channels and for the development of bifurcations from branches and channels.
 [S1063-7842(99)00604-2]

Single- and multichannel leader discharges over water surfaces similar to Lichtenberg figures have been studied by the author.[1–3] The Ohmic conductivity of water was found to cause a nonlinearity in the $R(t)C$-discharge circuit and a one-to-one correspondence between the channel lengths and the current passing through them, and that, as a whole, the discharge develops in a self-consistent manner. Given the multiplicity of elements in the structure of discharges of this type, the self consistency of their development must also have an effect on their structure. This paper is devoted to a study of the structure of leader discharges over water surfaces and its relationship to the dynamics of their development based on the earlier experimental work.[1,2]

EXPERIMENTAL CONDITIONS

The experimental setups of Refs. 1 and 2 are sketched in Fig. 1a and 1b. They consisted of a cell with water (*1*), a storage capacitor $C=0.1\mu$F (*2*), an anode in the form of a wire 0.0075 cm in diameter or a tip with that radius (*3*), and a cathode (*4*). A detailed description of the setup and diagnostic methods is given in the earlier papers. Both experiments used the same range of initial voltages on the capacitor ($U_0=3-6$ kV), conductivity of the water ($\simeq 1\times10^{-4}\Omega^{-1}\cdot$ cm^{-1}), length of the air gap between the anode and water surface (0.1–0.3 cm, depending on U_0), and depth of immersion of the cathode below the water surface ($\simeq$0.3 cm). Only the types of cathode were different. In Ref. 1 a one dimensional cathode consisting of a wire 8 cm long with a diameter of 0.16 cm was used (Fig. 1a) and in Ref. 2, a flat, circular cathode with a diameter of 7 cm (Fig. 1b). This resulted in a substantially different distribution of the electric field at the water's surface. In the first case a single-channel discharge develops and in the second, a multichannel discharge (Fig. 2). In both cases, however, the character of the $R(t)C$-circuits and their nonlinearity were determined by the resistance of the layer of water above the cathode and their electrical characteristics varied with time in a qualitatively similar fashion.[1,2]

After a voltage U_0 was applied to the anode, a pulsed corona developed in the air at its tip. As the corona lengthened and reached the water's surface, the current in the circuit rose and the discharge channel contracted and grew toward the surface of the water. The potential jump at the water's surface at the time the current channel reached the water initiated the formation of several radial leader channels in the plasma spot on the water. As they developed, the area of the spot increased, as well as that of the layer of water above the cathode through which the current flowed. This reduced the resistance of the layer and caused the current to rise further. As the current rose, and, therefore, the power dissipated in the circuit increased, the resistance of the discharge channel in the air and the resistance of the leader channels fell; this reduced the voltage drop across them and sustained the potentials at their tips, enabling them to move onward.[1] This, in turn, ensured a further increase in the current. Naturally, in the case of the one-dimensional cathode this occurred only for the channel that developed above it, except during the beginning of the discharge, when the conditions for all the channels were similar. This sort of one-to-one correspondence between the channel lengths l and the current i in them, which can be regarded as a positive feedback, was maintained until the potentials of the tip and the water below it were no longer equal owing to the rise in the voltage drop across the leader channel. At that time t_M, the leader stopped and its length and the current through it reached their peak values (l_M and i_M). From this time on, the

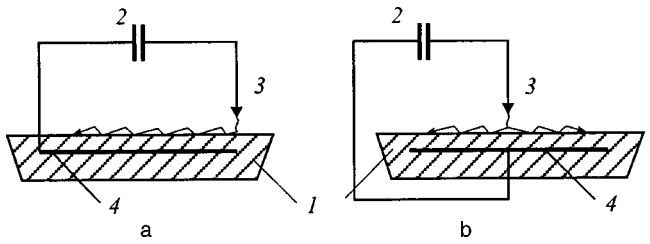

FIG. 1. A sketch of the experimental apparatus: (a) single-channel and (b) multichannel leader discharges.

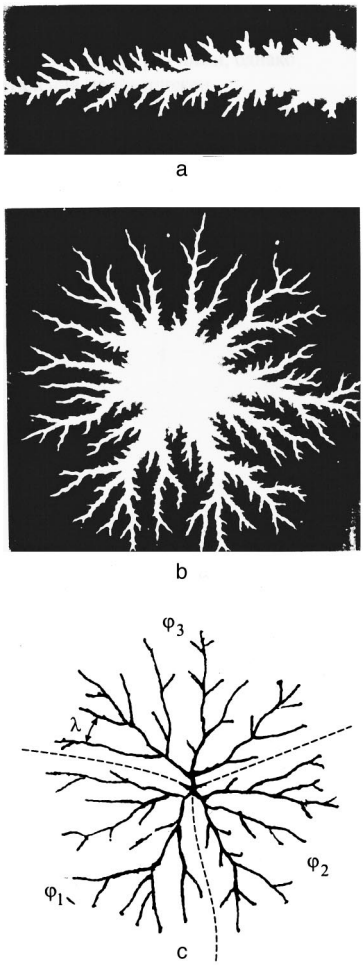

FIG. 2. Single-channel (a) and multichannel (b) leader discharges ($U_0=6$ kV) and the coarse structure (c) of a multichannel leader discharge from the photograph of (b). Scale 1:1.

power losses in the expanding channel exceeded the power released there. The disruption of the power balance in the leader channel caused the plasma to decay, especially in the tip, where the losses are greater. This caused a contraction of the channel length and accelerated the current drop because of an l, i-correspondence which set in as a negative feedback. Thus, the individual channels and the entire structure of the discharge as a whole develop only during the stage when the current is rising with positive feedback between the channel length and the current passing through it.[1,2]

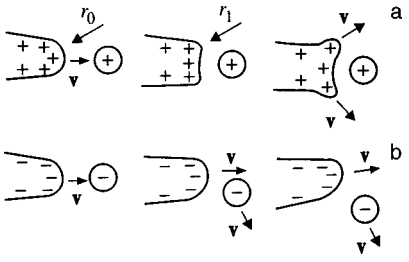

FIG. 3. A sketch of the interaction of a leader tip with the residual charge: leaders (a) moving toward the emitter and (b) moving toward the collector.

The final structure of a multichannel discharge consists of 3–6 initial channels, their branches, and their subbranches and those of the channels.[2] In accordance with an analysis of the dynamics of the channel development in Ref. 2, in this case two channels *1* and *2* (sectors φ_1 and φ_2 of Fig. 2b and 2c) appear synchronously and then channel *3* (sector φ_3), completing the formation of a stable initial electrostatic structure. Here the first two initial channels developed radially in one half plane and the third channel, in the other half plane. This is related to the axial symmetry of the anode field at the water's surface and its dominant role in the initial development of the channels. The first bifurcations of the channels are, respectively, $\varphi_1\simeq\varphi_2\simeq 80$ deg and $\varphi_3\simeq 140$ deg. However, the angles of the later bifurcations are close to 40–50 deg for all the channels, and the final structure of the branches of the first two initial channels occupies a sector $\varphi_1+\varphi_2>180$ deg. This all indicates that, with distance from the anode, the local field in front of the tips of the channels affects the discharge structure; this field is random and causes the differences in the channel development. Therefore, the discharge structure is formed by the motion of the initial channels and their branching as they move under the influence of the central and local electric fields.

Channel branching is fundamentally intrinsic to extended discharges in air, but its mechanism has apparently not been specially studied. Bazelyan and Razhanskii[4] have considered the branching of a streamer as the development, with time, of a second streamer on its side when the radial field strength exceeds 150 kV/cm and the effective potential, 10 kV. In our case, these magnitudes are not so high and photographs show that the branches of the final bifurcations have roughly the same lengths, which suggests that the branches develop at the same time. Thus, another mechanism for branching must be invoked. Schonland[5] considered the cause of the convolution and branching of discharge channels to be the space charge in front of them, but did not propose a specific mechanism.

The data of Refs. 1 and 2 indicate that the reason for leader branching is the splitting of its tip owing to flattening of its initially hemispherical shape with radius r_0 because of the interaction between its charge and the residual charge of the same sign in front of the tip which accumulates as the tip moves (Fig. 3a). The flattening of the tip causes regions with radius $r_1<r_0$ to appear at the channel edge near the generatrices of the channel where the field strength is higher than at the center of the tip along the channel axis. Since the velocity depends on the field strength, with the passage of time the central part of the front of the tip will slow down, while the peripheral regions will accelerate. Finally, this leads to bifurcation. Let us evaluate the feasibility of this mechanism by comparing the average time interval for leader branching with the calculated time required for a residual positive charge equal to the charge in the tip to develop ahead of the tip based on the proposed mechanism and the experimental conditions. The residual charge ahead of the tip is a consequence of the corona discharge from the tip and its magnitude should be determined by the corona current i_k and the branching time Δt. When the leader finishes developing, the potential $\varphi\simeq 2$ kV at its tip[1] is roughly the same as the igni-

tion (quenching) potential for corona in air and $i_k \simeq 1 \times 10^{-6}$ A.[6] The charge q_h at the tip, however, is determined by its potential and capacitance C_h with respect to the cathode. For a diameter of the last centimeter of the leader length l_k close to the diameter of the tip, $\varnothing_h \simeq 1 \times 10^{-2}$ cm, a water layer of thickness $d \simeq 0.3$ cm, and a dielectric permittivity for water of $\varepsilon \simeq 80$, the capacitance of this centimeter-long leader is

$$C_h \simeq \frac{\varepsilon \varnothing_h \cdot l_k}{4\pi \cdot d} \simeq 2 \times 10^{-13}\,\text{F}.$$

Since essentially only the tip, i.e., a segment of the leader of length $\simeq 10^{-2}$ cm, interacts with the residual charge, we have $C_h \simeq 2 \times 10^{-15}$ F. For $\varphi_h \simeq 2$ kV, its charge and the branching time will be

$$q_h = C_h \varphi_h \simeq 4 \times 10^{-12}\,\text{C}, \quad \Delta t = \frac{q_h}{i_k} \simeq 4 \times 10^{-6}\,\text{s}.$$

In the initial stage of the development of the leader, $\varphi_h \simeq 6$ kV and, since $q \propto \varphi_0$ while $i_k \propto \varphi_0^2$,[6] in this case $\Delta t \simeq 1 \times 10^{-6}$ s. At the same time, judging from Fig. 2a, the tip split 5–6 times in the last centimeter of the leader and 3–4 times in the first centimeter while they traveled, respectively, for $\simeq 10$ and $\simeq 1\,\mu$s.[1] Thus, the average branching times at the end and at the beginning of leader development were $\simeq 2$ and $\simeq 0.3\,\mu$s, respectively, which are of the same orders of magnitude as the times calculated above.

Similarly, in the case of a negative leader moving toward the anode, a negative residual charge will develop ahead of the tip and inhibit its motion (Fig. 3b). However, the lower mass of electrons will promote a higher transverse (relative to the trajectory of the leader) lability of this discharge. And in this case, first of all, the tip will split up faster and, second, the deflection angles of its trajectory (without splitting) will be smaller. Smaller amounts of branching and convolution of the channels of leaders moving toward the anode have actually been observed.[7]

As noted above, one consequence of the branching mechanism proposed here is fluctuations in the axial velocity of the leader. This has also been observed experimentally[4] and confirms the existence of a branching mechanism associated with splitting of the leader tip. During splitting without neighboring charges, parts of the tip will tend to move apart in diametrically opposite directions, as suggested by the first bifurcation angle of $\varphi_3 \simeq 140$ deg. Later when the tips split, the resulting branches already have neighbors so their electrostatic interaction reduces the bifurcation angles to 40–50 deg. In addition, the electrostatic interaction of the tips of three successive branches of a single channel, which form a fan (Fig. 2), makes the central branch tend to move rectilinearly, so it has better conditions for further development. At the same time the side branches will inevitably compete with the branches of neighboring channels and have a higher probability of ceasing. This determines the symmetry of the structure of a branch with respect to its channel and the regularity of the discharge structure along the perimeter. Under actual conditions, the splitting of a tip and its trajectory are also affected by external random factors (films, microscopic particles on the water's surface) which disrupt the regularity of the discharge structure. Thus, a set of cause and effect factors, including the flattening of a tip and the dependence of the velocity of its front on its radius, as well as some random interactions of the tips, determine the stochasticity of the process by which the discharge structure is formed.

It has been found[2] that the bifurcations from the channels and branches develop normal to them and with a delay relative to the movement of the tips. Given that $t_M \simeq 20\,\mu$s in a single-channel discharge and bifurcations are present along almost the entire channel length (Fig. 2a), while $t_M \simeq 4\,\mu$s in a multichannel discharge, while bifurcation is observed only at the base of the channels (Fig. 2b), we conclude that the time constant for formation of the bifurcations is $\leqslant 4\,\mu$s under these conditions. The fact that bifurcations develop in the depth of the structure and normal to its elements indicates that their development is entirely controlled by the local field and not by the anode field.

As can be seen from Fig. 2a and 2b, bifurcations develop against a background of a diffuse luminosity along the channels, i.e., in the boundary layer of a nonequilibrium plasma. Their appearance and regularity ($\Delta l \simeq 5 \times 10^{-2}$ cm) can be ascribed to the fact that they are initiated by bunches of ions in the longitudinal ion acoustic oscillations excited by the fluctuating motion of the leader tips. As noted above, the first centimeter of the tip channel length is covered over a time $\Delta t \lesssim 1\,\mu$s, and it splits up 3–4 times ($n = 3$–4), i.e., the frequency of the fluctuations in the tip velocity is $\omega = (2\pi n)/(\Delta t) \simeq 2 \times 10^7\ \text{s}^{-1}$. Assuming that the ion temperature is less than the electron temperature, i.e., $T_i < T_e \simeq 0.3$ eV,[4] we find the ion sound speed and wavelength of the longitudinal ion oscillations to be[8]

$$C_h = \left(\frac{T_e + \frac{5}{3} T_i}{M_i} \right) \simeq 2 \times 10^5\,\text{cm/s},$$

and

$$\lambda_i = \frac{C_i \cdot 2\pi}{\omega} \simeq 6 \times 10^{-2}\,\text{cm}.$$

This wavelength of the ion oscillations, i.e., the interval between the bunches in the ion density, is fully consistent with the interval between the bifurcations and thereby confirms the proposed bifurcation mechanism.

Judging from an enlarged image of the middle portion of a leader (Fig. 2a) (better focused), the bifurcation also has a structure but it cannot be analyzed because of the insufficient resolution in this picture. It may be assumed, however, that the structure of the bifurcations is also the result of splitting of leader tips. Thus, the structure of the discharge as a whole consists of two hierarchical stages, whose development is spaced in time with different mechanisms for the appearance of the steps and for formation of their elements.

Besides the electrostatic factors considered here, a current competition among the leader branches affects the discharge structure. The phenomenon of current competition among the channels of multichannel discharges in dense gases is little studied. In its general form, the scheme of this

competition is universal: a random initial predominance of one of the channels allows it to receive more energy, which, as it is dissipated in the channel, ensures a higher conductivity and rate of development and, therefore, a subsequent advantage in consuming energy from the source. The simultaneous reduction in the latter time makes it more difficult for other channels to develop. A scheme of this sort including an l, i-correspondence, however, provides much more detail and makes it possible to relate the final preeminence of a leader or branch in the picture (large length and sector occupied by the branches) to a random advance in time or a random predominance in its site of origin.

It is clear in Figs. 2b and 2c that in the initial stage of development of leaders, where $r<0.5$ cm, the tips are able to spread out and the current competition between the leader branches is minimal. In the range $r\simeq 0.5$–1 cm, the side branches of all the leaders are close in perimeter to within a distance $\lambda\simeq 0.1$ cm, the current competition among them increases, and the electrostatic interaction, in turn, keeps the tip from splitting. Only when the radius has increased to 1.2 cm and the corresponding increase in the perimeter of the structure and reduction in the number of branches have taken place owing to current competition does the distance between the tips increase and they again begin to split. Here the branches which overtake their neighbors, i.e., win in the current competition (l, i-correspondence), gain the opportunity to spread further and finally suppress the lagging branches. This combination of electrostatic and current interactions among the leaders leads to the periodic formation of large elements in the discharge structure and it becomes hierarchical in terms of both its perimeter and its radius. Furthermore, both these interactions couple the development of each leader to the development of the entire discharge structure, which implies that the discharge is self consistent.

The self-consistent development and hierarchical nature of the structure suggest that it must be fractal. It has been found[9] that the structure of a leader discharge over a dielectric surface (Lichtenberg figure) has a fractal dimensionality $D\simeq 1.7$. This means that the overall length of all the elements in a structure within a circle of specified radius is $L(r)\propto r^D$. The fractal dimensionality is defined as the slope on a log–log plot of $L(r)$. A computer simulation of fractals of this sort (referred to as Laplacian because of the corresponding field distribution) gives a dimensionality of $D=1.75\pm 0.02$ assuming that the probability of bond breaking (a square lattice constant) is proportional to the local field strength.[9,10]

It is rather difficult to determine $L(r)$ with the required accuracy from photographs of the discharge. However, since the number (of bonds) of the branches intersecting a circle of radius r is

$$n(r)=\frac{dL(r)}{dr}\propto r^{D-1},$$ [11]

the fractal dimensionality of the discharge structure can be determined using this formula. For this purpose, an annular grid was superimposed on Fig. 2b and the number of branches and channels intersecting a circle of given radius was determined using it. Figure 4a shows that the number of

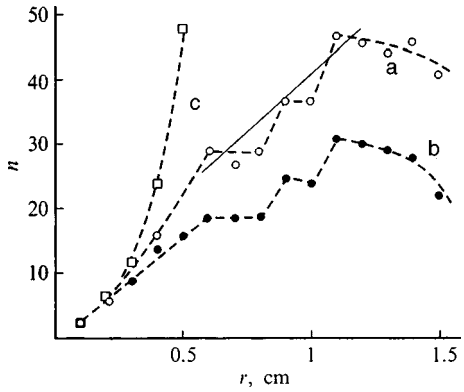

FIG. 4. the number of structural elements of a multichannel leader discharge as a function of radius: (a) actual discharge, (b) coarse structure, (c) doubling the number of branches over a 0.1 cm channel length; $U_0=4$ kV.

intersections of elements of the structure as a function of radius for $U_0=4$ kV, beginning with $r=0.5$ cm, has a fluctuating character. For $U_0=6$ kV, the fluctuations began at $r=0.9$ cm. The fluctuations in the graph remained when the number of elements was counted by different observers, although the number of elements that were recorded varied because of their different levels of perception. Furthermore, the character of the graph does not change if we consider only branches with lengths (wavelengths) greater than 0.2 cm (the structure is coarsened) (Fig. 4b). In the initial stage, up to $r<0.3$ cm for $U_0=4$ kV, the actual number of branches (Fig. 4 a) and their possible number upon doubling over 0.1 cm (Fig. 4c) differ little. This is because along the perimeter of the structure the tips of the channels are quite separated; they essentially develop and bifurcate under the influence of only the central field under l, i-correspondence conditions and when only a positive feedback exists. As the distance between the tips decreases, their splitting begins to be slowed down by the charge interaction, which serves as negative feedback. The existence of positive and negative feedback in the system inevitably causes oscillations during its development (the fluctuations in Fig. 4).

These features of the development of the structure are more clearly visible in a plot of the average distance between branches, $\bar{\lambda}=2\pi r/n(r)$, as a function of radius (Fig. 5). This figure shows that, for $U_0=4$ kV in the interval $0.5<r<1.2$ cm a separation $\bar{\lambda}=0.16$ cm is achieved and maintained, with oscillations, that is close to the minimum distance between the elements of the structure. Similarly, for

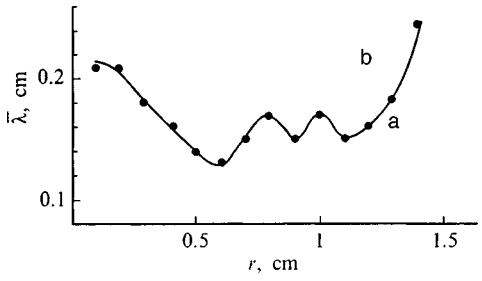

FIG. 5. The average distance between branches as a function of radius ($U_0=4$ kV).

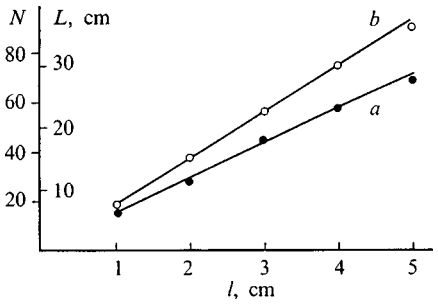

FIG. 6. The total length of structural elements of the discharge (a) and the number of bifurcations in the discharge structure (b) as functions of the channel length of a single-channel leader discharge $U_0=6$ kV.

$U_0=6$ kV, in the interval $0.9<r<2$ cm, $\bar{\lambda}=0.2$ cm, i.e, a maximal density of filling the space with elements of the structure is maintained. Linear $n(r)$ dependences which average these fluctuations (the smooth curve of Fig. 4a) were used to calculate the fractal dimensionality. In all cases, for $U_0=4$ and 6 kV and the three or four initial channels, $D=1.85\pm0.05$.

The fractal dimensionality of the structure of a single-channel discharge (Fig. 2a) was determined using two relations: $L(l)\propto l^D$ and $N(l)\propto l^D$. In the first case, $L(l)$ is the length of all elements of the structure over a length l (Fig. 6a), which was measured with a curvometer on a three-times enlarged photograph. In the second case, $N(l)$, the number of bifurcations in the structure of a channel in length l (Fig. 6b), was determined from the same photograph. In a channel segment of length 4 cm from the anode, where $L(l)$ and $N(l)$ are linear, the corresponding values of the fractal dimensionality of the structure were $D=0.96\pm0.05$ and 0.97 ± 0.03, respectively.

The difference between these values for the fractal dimensionality of a leader multichannel discharge, $D=1.85$, and for a single-channel discharge, $D=0.96$, is related to the two dimensional and quasi-one-dime nsional characters, respectively, of the electric field distributions in these cases, as well as to the relationship between the deterministic and random factors in the development discharge structure. The longer time $t_M\simeq20\ \mu$s for a single-channel discharge to develop compared to the $t_M\simeq4\ \mu$s for a multichannel discharge offers a greater possibility for randomness to show up. This circumstance also partially explains the difference between the value of $D=1.85$ for a multichannel discharge over a water surface and $D=1.7$ over a solid dielectric,[9] since the characteristic time for a discharge to develop in the latter case is $\simeq1.0\ \mu$s. In addition, the difference in the values of D is explained by the large degree of filling of the space with structural elements of the discharge in our case, since the elements developed, not only along the perimeter in the ''growth'' zone, as in a solid dielectric,[9,12] but also in the inner regions of the structure owing to the Ohmic conductivity of water and the flow of a current to the cathode over the entire area of the water layer above it within the confines of the structure.

The results of Refs. 1–3 and this paper yield the following conclusions:

1. The structure of a leader discharge is formed by two hierarchical stages. The first involves the simultaneous appearance of the initial channels and their subsequent branching, mainly in the field of the anode, and the next step, the development of bifurcations off the channels and branches in the local fields of the structure.

2. Both stages proceed through the movement and splitting of the tips. The motion of a tip is determined by its charge and the field strength, while splitting is related to the flattening of its shape as it moves.

3. The increase in the number of channels of the branches owing to splitting of their tips introduces a randomness in the development of the discharge and causes competition and selection of channels owing to their charge interaction and the l, i-correspondence.

4. The selection of channels during the self-consistent development ensures a hierarchical and fractal structure of the leader discharge.

5. Side branches and bifurcations are elements of the hierarchical structure of a leader but they are also inherent factors in its development which create the conductivity of the leader channel and thereby maintain the potential at the tips.

6. The Ohmic conductivity of water is responsible for the nonlinear development of the channel through the l, i-correspondence, which sets in as a positive feedback for $di/dt>0$ and a negative feedback for $di/dt<0$.

These results also offer the possibility of comparing leader discharges over water (dielectric) surfaces with atmospheric leaders in a most general way. Although the elementary processes in their plasma fronts are the same, the formation of macroscopic structures and the macroscopic processes in an atmospheric leader take place in a higher field and, therefore, have a more marked deterministic character (less branching). On the other hand, the delayed kinetics in the low longitudinal field strength of a leader along a water surface opens up greater possibilities for random processes and the leader development becomes more stochastic. This should lead to a difference in the fractal dimensionalities of their structures. On the other hand, the importance of transverse structural elements (lateral branches and bifurcations) in the development of leaders on water surfaces indicates that they must be taken into account in developing a model for atmospheric leaders.

In conclusion, it might be added that using water, rather than a solid dielectric, to study leader discharges eliminates the need to replace the samples, since the properties of the water surface do not change. The ease of changing the conductivity of water makes it possible to vary the dissipative losses in a controlled way, which is important for establishing the relationship between the losses and the fractal dimensionality of a leader discharge structure.

[1] V. P. Belosheev, Zh. Tekh. Fiz. **66**(8), 50 (1996) [Tech. Phys. **41**, 773 (1996)].
[2] V. P. Belosheev, Zh. Tekh. Fiz. **68**(7), 44 (1998) [Tech. Phys. **43**, 783 (1998)].
[3] V. P. Belosheev, Zh. Tekh. Fiz. **68**(11), 63 (1998) [Tech. Phys. **43**, 1329 (1998)].

[4] É. M. Bazelyan and I. M. Razhanskiĭ, *Spark Discharges in Air* [in Russian], Nauka, Novosibirsk (1988).
[5] B. F. J. Schonland, *The Flight of Thunderbolts*, 2nd ed. [Clarendon Press, Oxford (1964); Gidrometeoizdat, Moscow (1970), 83 pp.].
[6] Yu. P. Raizer, *Gas Discharge Physics* [Springer-Verlag, New York (1991); Nauka, Moscow (1987), 505 pp.].
[7] I. S. Stekol'nikov, *The Nature of Long Sparks* [in Russian], Izd. AN SSSR (1967), 90 pp.
[8] L. A. Artsimovich and R. Z. Sagdeev, Plasma Physics for Physicists [in Russian], Atomizdat, Moscow (1979), 64 pp.
[9] J. Nittmann, G. Daccord, and H. Stanley, Nature (London) **314**, 141 (1985).
[10] L. Niemeyer, L. Pietronero, and H. J. Wiesmann, Phys. Rev. Lett. **52**, 1033 (1984).
[11] H. J. Wiesmann and L. Pietronero, *Fractals in Physics*, Mir, Moscow (1988), 211 pp.
[12] L. Pietronero, K. Ewerts, and H. J. Wiesmann (eds.), *Fractals in Physics* [North-Holland, Amsterdam (1986); Mir, Moscow (1988), 211 pp.].

Translated by D. H. McNeill

Generation of pulsed pressure in a liquid by means of a metallic plasma and measurement of its characteristics

V. P. Kortkhondzhiya, M. O. Mdivnishvili, and M. I. Taktakishvili

Institute of Physics, Academy of Sciences of Georgia, 380077 Tbilisi, Georgia
(Submitted November 20, 1997; resubmitted June 1, 1998)
Zh. Tekh. Fiz. **69**, 41–43 (April 1999)

Certain characteristics of pulsed pressure generated in a liquid by means of a metallic plasma are investigated experimentally. The problem of increasing the duration of the pressure pulse by plasma-chemical reaction is discussed.
[S1063-7842(99)00704-7]

Pulsed pressure can be created in a liquid by many techniques:[1–4] by means of explosives, spark discharge, the electrohydraulic effect, etc. Of particular interest is the explosion of a wire in a liquid, because: a) the technique does not require high voltages, so that pulsed pressures can be produced in a cavity of comparatively small dimensions; b) a spatial profile is easily imparted to the pressure; c) the technique is inexpensive and readily implemented.

The objective of the present study has been to investigate certain characteristics of pulsed pressure generated in a liquid by a metallic plasma with a capacitor bank embedded in the exploding wire and operating at high energies. We have investigated two regimes of this process: 1) the production of pulsed pressure by the explosion of a wire whose material does not enter into plasma-chemical reaction with the liquid; 2) the production of pulsed pressure when the material of the exploding wire enters into plasma-chemical reaction with the liquid in which explosion takes place.

A pulsed pressure is created in a liquid by an exploding wire as follows: A wire of copper, aluminum, or other material is strung across a cavity filled with a slightly compressible, neutral liquid by means of insulators.

A high-current electrical pulse is delivered to the wire, causing the wire to melt and evaporate and thus generating a metallic plasma, which is characterized[5] by high temperature, a large number of particles $\sim 10^{20}\,\mathrm{cm}^{-3}$, and high pressure. This pressure is transmitted from the metallic plasma to the cavity walls through the liquid occupying the cavity.

Figure 1a shows an oscillogram of the pulsed discharge current for the detonation of a copper wire in water. Clearly, the energy of the capacitor bank is spent in exploding the copper wire in a time of the order of 0.2 ms. Certain conditions, to be discussed below, must be met in this case.

The pulsed pressure generated by the exploding wire is measured by two methods, one using a crusher gauge[6] and the other by means of a piezoelectric pickup.[7,8] The crusher gauge comprises a metal cylinder moving in a metal branch piece soldered to a tube. The measurement of pressure by means of a crusher gauge operates on the principle that the deformation imparted to the gauge by the pressure is proportional to this pressure. The crusher gauge must be calibrated before it can be used to measure pulsed pressure. It is calibrated by a static pressure for several crusher gauges of a single batch fabricated from the same blank. This step is necessary in order to reduce the subsequent measurement error. In the calibration operation a graph of the absolute change $l = l_0 - l_1$ or the relative change in length of the crusher gauge under the influence of a static pressure of known magnitude is plotted. Here l_0 is the initial length of the gauge, and l_1 is its length after deformation. This graph is subsequently used to determine the amplitude of the pulsed pressure from the deformation of the crusher gauges.

A piezoelectric pickup can be used to trace the time variation of the pulsed pressure. Figure 1b shows a pulsed pressure oscillogram obtained in the detonation of a copper wire of diameter 1.5 mm and length 300 mm in a water-filled tube of diameter 500 mm. After the first peak, which corresponds to detonation of the wire, several peaks of smaller amplitude are observed. They are associated with multiple reflections of waves from the vessel walls and are well resolved by the piezoelectric pickup. The duration of the pressure peak created by the detonation of the copper wire is of the order of 30 μs. The piezoelectric pickup is calibrated by means of a crusher gauge, which records the amplitude of the pulsed pressure. The variation of the output voltage pulse of the pickup as a result of the pulsed pressure as a function of the latter in the piezoelectric pickup structure used by us is

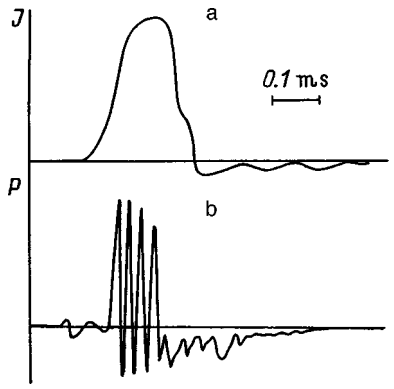

FIG. 1. Detonation of a copper wire in water. The energy stored in the capacitor bank is 22.5 kJ, the wire diameter is 1 mm, and the length of the wire is 30 cm: (a) oscillogram of the discharge current; (b) oscillogram of the pressure (the pressure maximum is $\sim 10^8$ Pa).

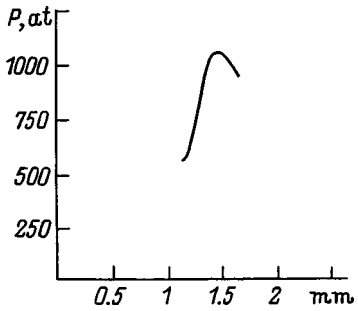

FIG. 2. Dependence of the pulsed pressure amplitude on the diameter of a wire of length 30 cm. The material of the wire is copper, and the energy of the capacitor bank is 22.5 kJ.

linear up to $7\times10^7\ N/m^2$. The structure of the pickup comprises a ring in which is packed a soft metal, a piezoelectric element, and more soft metal in sequence. This structure safeguards the piezoelectric element against fracture in the measurement of pulsed pressure in a slightly compressible liquid.

Experiments on the generation of pulsed pressure by means of an exploding wire have demonstrated the existence of optimal conditions conducive to the development of maximum pressure. Figure 2 shows the dependence of the pulsed pressure obtained in the detonation of a copper wire on the diameter of an exploding wire of specified length for a specified energy of the capacitor bank. It is evident that the maximum pulsed pressure in the detonation of a wire of length 300 mm for a fixed energy of the capacitor bank equal to 22.5 kJ is obtained when the copper wire has a diameter of 1.4 mm. Consequently, the parameters of the given experiment for a wire of diameter 1.4 mm are optimal for maximizing the pulsed pressure.

The conditions for maximizing the rate of energy release in the discharge gap in the underwater electrical detonation of wires have been determined in Refs. 9 and 10. An equation is given there for calculating the optimum exploding wire diameter. An estimate of the diameter of a copper wire from this equation under the optimal conditions of our experiment gives a result consistent with the experimentally observed value within 10% error limits.

It is important to note the specific characteristics of the detonation of wires of certain materials in a liquid when the generated metallic plasma enters into plasma-chemical reaction with the liquid. The plasma-chemical reaction can be written in the general form

$$n\cdot M+m\cdot L\rightarrow n\cdot K+m\cdot G+Q,$$

i.e., the interaction of n metal (M) atoms with m liquid (L) molecules results in the formation of n molecules of a compound K and G gas atoms with the release of energy Q from the exothermic reaction. This reaction is important for two reasons. First, the generated gas G develops an additional pressure, which is added to the pressure developed by the metallic plasma. Second, the attendant plasma-chemical reaction bears a certain resemblance to the detonation of a high explosive, and the duration of the pressure pulse is longer than in the detonation of a wire whose plasma product does not enter into reaction with the liquid.

These considerations are illustrated by the detonation of aluminum, zirconium,[11] and titanium in water. In particular, when an aluminum wire is detonated, the aluminum vapor attains a temperature of 3000 K, and the following reaction can take place:

$$2\,Al+3H_2O\rightarrow Al_2O_3+6H,$$

$$Al+H_2O\rightarrow Al(OH)_3+3H.$$

As a result of these reactions, three hydrogen atoms are associated with a single aluminum atom, and heat is released. The hydrogen generated in the process develops a pressure that is added to the pressure of the aluminum plasma. Moreover, for the reason stated above, the duration of the pressure pulse is greater than in the detonation of a copper wire, whose plasma does not enter into chemical reaction with water. The difference is evident from the following table:

Wire material	Cu	Al
Liquid occupying the cavity	H_2O	H_2O
cavity diameter, mm	50	50
Wire diameter, mm	1	1
Wire length, mm	230	230
Pressure amplitude, $\times10^5$ N/m^2	560	1100
Pressure pulse duration, ms	0.18	1.0

The detonation of the wire is accompanied by strong ultraviolet radiation, which can ionize the resulting vapor-gas mixture. This effect can lead to electrical breakdown of the mixture, where the greater the residual voltage across the

FIG. 3. Oscillograms of the discharge current (upper traces) and residual voltage (lower traces) in the detonation of an aluminum wire of diameter 1 mm and length 30 cm in water: (a) detonation at a discharge voltage of 4 kV; (b) detonation at 5 kV.

discharge gap after detonation of the wire, the earlier will bethe onset of electrical breakdown of the vapor-gas mixture. Oscillograms of the current and voltage generated by the detonation of an aluminum wire are given in Fig. 3. Figure 3a corresponds to the detonation of an aluminum wire of diameter 1 mm and length 100 mm at a discharge voltage of 4 kV. The voltage remaining in the discharge gap after detonation of the wire is not sufficient for electrical breakdown of the vapor-gas mixture. When the voltage is increased, breakdown of the vapor-gas mixture takes place after a certain time lapse attributable to a drop in the pressure to a level at which breakdown can occur. It is evident from the oscillogram in Fig. 3b, which was obtained at a voltage of 5 kV, that the discharge current pulse at the left end of the upper trace and the pulse associated with breakdown of the vapor-gas mixture have almost merged, while the voltage oscillogram represented by the lower trace indicates that all the energy stored in the capacitor bank is consumed in detonating the aluminum wire of diameter 1 mm and length 100 mm and in electrical breakdown of the vapor-gas mixture. The latter event, in turn, generates a second pressure pulse. It is important to note that because the water is heated to a high temperature and partially dissociates, the excess of atomic hydrogen produced by the plasma-chemical reaction and the oxygen subsequently explode as well.

We summarize the work in a few concluding remarks. Short (of duration $25-30\ \mu$s) pressure pulses can be generated by creating a pulsed pressure in a liquid by means of a metallic plasma.

If the metallic plasma enters into plasma-chemical reaction with the liquid, the generation of a gas by the plasma-chemical reaction can produce a pulsed pressure with at least twice the pressure amplitude obtained in the detonation of a wire in the liquid when plasma-chemical reaction does not occur.

If the metallic plasma enters into plasma-chemical reaction with the liquid, the duration of the pressure pulse is several times the duration obtained without plasma-chemical reaction.

In the case of electrical discharge in water 50% of the energy is spent in the dissociation of water,[1] and only the remainder is left for generating pulsed pressure. According to our data, when pulsed pressure is generated by means of a metallic plasma under optimal experimental conditions, almost all the energy of the capacitor bank, exclusive of the part spent in heating the conducting wires, actuating the spark gap, etc., goes for the generation of pulsed pressure.

The authors are grateful to N. A. Kervalishvili, G. I. Suramlishivili, M. K. Lortkipanidze, and G. P. Tsikarishvili for their interest and valuable discussions.

[1] K. A. Naugol'nykh and N. A. Roĭ, *Electrical Discharges in Water* [in Russian], Nauka, Moscow, 1971, 259 pp.
[2] S. A. Dorofeev, *The Electrohydraulic Effect and Its Applications* [in Russian], Naukova Dumka, Kiev, 1961, 311 pp.
[3] V. M. Demina and A. A. Shkatov, *Physical Principles of Electrical Detonation* [in Russian], Naukova Dumka, Kiev, 1983, 352 pp.
[4] L. A. Yutkin, *The Electrohydraulic Effect and Its Industrial Application* [in Russian], Mashinostroenie, Leningrad, 1986.
[5] *Plasma Accelerators*, edited by L. A. Artsimovich *et al.* [in Russian], Mashinostroenie, Moscow, 1973; *Physics and Application of Plasma Accelerators*, edited by A. I. Morozov [in Russian], Minsk, 1974.
[6] W. Golke, *Physics of Short-Time Processes*, Vol. 2 (Mir, Moscow, 1971), p. 69.
[7] V. S. Komel'kov and V. I. Sinitsyn, *Plasma Physics and the Problem of Controlled Thermonuclear Research* [in Russian], Izd. AN SSSR, Moscow, 1958.
[8] V. K. Kedrinskiĭ and R. I. Soloukhin, Prikl. Mekh. Tekh. Fiz., No. 1, 27 (1961).
[9] *Physics Encyclopedic Dictionary* [in Russian], Sov. Éntsiklopediya, Moscow, 1983.
[10] E V. Krivitskiĭ, *Dynamics of Electroexplosion in a Liquid* [in Russian], Naukova Dumka, Kiev, 1986, 401 pp.
[11] *Electrical Detonation of Wires*, edited by A. A. Rukhadze and I. S. Shpigel' (Mir, Moscow, 1965), pp. 239–259.

Translated by James S. Wood

Construction and study of a magnetoactive plasma ion source

M. A. Krasnogolovets

Kharkov Technical University for Radio Electronics, 310108 Kharkov, Ukraine
(Submitted December 15, 1998)
Zh. Tekh. Fiz. **69**, 44–47 (April 1999)

The interaction of intense electron beams with plasmas in a nonuniform magnetic field is studied for the purpose of obtaining a magnetoactive plasma ion source. The resulting experimental data are used to trace the dynamics of the changes in the basic parameters of the source, both during the stage where it interacts with the beam and in the cooling stage. It is found that the charged particles are contained for a long time and that the efficiency of energy transfer from the beam to the source is high. This source is intended to be used for shaping and accelerating multiampere ion beams. © *1999 American Institute of Physics.* [S1063-7842(99)00804-1]

Intense ion fluxes are necessary in a number of areas of science and technology, in particular for creating short-wavelength coherent radiation sources.[1–4] At present, the theory of ultraviolet and x-ray lasers and their practical realization are greatly impeded by a lack of experimental information on the cross sections for elementary events such as the cross section σ_{in} for n-fold ionization of an atom, the lifetime τ_{in} of multiply charged ions in excited states, the n-fold ionization potential E_{in}, and so on. There are no experimental data on elementary events for the collision cross sections of electrons with excited atoms, even those with a simple electronic structure (He, Li, Na). On this topic an experimental study of the interaction of a high power electron beam with a plasma has been undertaken in order to obtain an ion source with which a study of these parameters has been begun. These studies have led to rational choices of the electron beam current I, the beam acceleration voltage U, the working gas pressure P, the configuration and strength of the magnetic field B, and the geometric dimensions r and l of the system required to obtain optimal parameters of the ion source.

In various experiments with beam–plasma discharges it has been found that for relatively low neutral gas temperatures there is a group of ''superhot'' electrons with a temperature of about 10^9 K.[5,6] As a rule, there has been agreement that the heating of the plasma electrons is related to the excitation of slow waves with $V_\varphi < c$.[1–8] In the experiments various ways of transferring energy from an electron beam to the plasma electrons have been pursued and at present there is no single point of view regarding the mechanism of energy transfer. This is because of the complexity and inherent nonlinearity of the phenomena, the large influence of the initial and boundary conditions, and a lack of reliable experimental data.

We shall examine the most important discharge characteristics which can be used to elucidate its basic behavior. Not enough attention has been paid to studies of the ionization of the neutral gas and the buildup of charged particles in the source to values $n \gg n_0$, where n_0 is the initial density of the neutral gas. In addition, these processes often control the macroscopic behavior of the beam–plasma interaction as a whole and are especially important for creating a plasma ion source.

The buildup of charged particles in the active region can be described by a system of balance equations for the electrons and ions in a beam–plasma discharge:[9]

$$\frac{\delta n_e}{\delta t} = \frac{n_e}{\tau_e} + I, \quad \frac{\delta n_i}{\delta t} = \frac{n_i}{\tau_i} + n_e\left(n_{01}k_0 + n_{02}\frac{V_1}{V_0}k_1\right). \qquad (1)$$

Here n_e, n_i, n_{01}, and n_{02} are the densities of electrons, ions, atoms, and molecular ions, respectively; V_1/V_0 is the ratio of the volume of the outer layer of the discharge, into which molecules penetrate, to the total volume; k_0 and k_1 are the probabilities of ionizing the atoms and of dissociating the molecular ions; I is the rate at which ions enter a unit volume of the discharge; and, τ_e and τ_i are the electron and ion lifetimes in the discharge. The greatest theoretical difficulty in solving Eq. (1) is to determine the electron lifetime τ_e in a system with nonuniform magnetic fields. We can write

$$\tau_e^{-1} = \tau_t^{-1} + \tau_k^{-1} + \tau_r^{-1} + \tau_c^{-1} + \tau_d^{-1}, \qquad (2)$$

where τ_t is the loss time through thermal conduction, τ_r is that through ionization, τ_c is that through loss of electrons along the axis of the system, and τ_d is that through radial loss of electrons because of turbulent diffusion.

All these times, as well as the probabilities of ionization and dissociation, depend to a substantial degree on the density of the electrons and their energies. In addition, in each stage of the discharge the effect of one or the other loss time in Eq. (2) may be utterly different. Thus the system of Eqs. (1) is essentially nonlinear. For that reason, the character of the ionization and charged particle buildup in the plasma source during the beam–plasma interaction is to be determined experimentally.

An electron beam is directed into a space filled with the working gas at a pressure of $P = 10^{-4}$ Torr. The beam was shaped and accelerated using an electron injector whose emitter was a lanthanum hexaboride wafer with a usable area of $S = 1.5$ cm^2. The maximum beam current was $I = 25$ A with an accelerating voltage $U = 30$ kV. The pulse duration was $\tau = 2$ ms. After the electron beam had passed through

the interaction region, it was dumped at a grounded collector. The beam–plasma interaction took place in an axially symmetric, stationary, nonuniform magnetic field with a mirror configuration. The mirror ratio was $\Phi = B_0/B_m = 1/4$, where $B_0 = 0.09$ T is the magnetic induction at the center of the system and B_m is that in the mirror region.

The stainless steel vacuum vessel had a diameter $d = 45$ cm in its central portion and the distance between the mirrors of the magnetic trap was $L = 40$ cm.

The energy of the electron beam was measured by parallel resistive and capacitive voltage dividers with the same attenuation and the current, by a movable collector in the high vacuum region where there were no collective effects owing to the interaction of the beam and plasma. The integrated visible luminosity of a small segment of the plasma ($r_0 = 0.5$ cm) was recorded using a photomultiplier with a collimated set of diaphragms. The electron density was estimated from the cutoff of microwave probe signals at wavelengths $\lambda = 0.8$ and 0.4 cm. The hard bremsstrahlung generated by the fast electrons was measured using scintillation detectors, as well as with a pinhole camera. The energy of the orbital motion of the charged particles per unit volume was determined with a diamagnetic probe consisting of a solenoid shielded from charged particle bombardment and electromagnetic pickup.

The energy spectra of the electrons and ions for energies $E \leqslant 10$ keV were analyzed using a multigrid probe. The change in the pressure of the working gas during the interaction between the beam and gas was measured with a pressure probe located 4 cm from the discharge boundary.

The simultaneous changes in the set of basic parameters of the ion source made it possible to determine the main processes responsible for forming the hot electron plasma, ionizing the working gas, and the buildup of charged particles in the system. Figure 1 shows a series of oscilloscope traces with a scan duration of 3 ms which can provide some idea of the variations in the major macroscopic characteristics of the source. Trace *a* characterizes the change in the voltage accelerating the electron burst.

The formation of an ion source with an electron temperature much greater than the ion and neutral gas temperatures, i.e., $T_e/T_i \gg T_e/T_0 \gg 1$ can be described in the following way. At first the electron beam ionizes the neutral gas through direct collisions with the gas molecules and after 30 μs a plasma with $n \geqslant n_{\text{beam}}$ is formed. Plasma oscillations with a growth rate $\gamma = (n_{\text{beam}}/n)^{1/3}\omega_{0e}$, where ω_{0e} is the plasma frequency. In these experiments, the characteristic length for pumping these oscillations is $L_1 \approx V_e/\gamma$ (where V_e is the beam velocity) or about ten centimeters, i.e., the conditions were favorable for efficient transfer of the energy of the electron beam to the gas. Nonlinear oscillations limited the growth in amplitude of the oscillations. As they interacted with these oscillations, the electrons in the source gained enough energy to ionize the working gas efficiently. At the same time,[10] electromagnetic noise was emitted from the system. Wave–particle interactions in the source caused diffusion of the charged particles. The average diffusion velocity $V_d = 1.2 \times 10^4$ cm/s after 700 μs. The ratio $r/R \ll 1$ (where r and R are the radii of the beam and ion source, respectively).

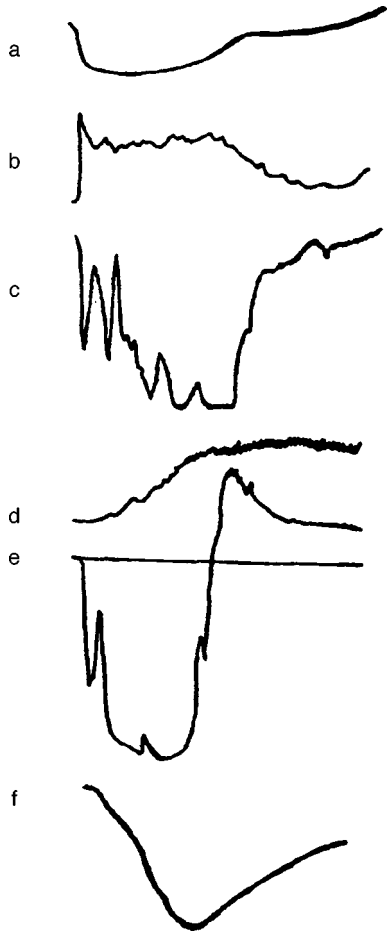

FIG. 1. Oscilloscope traces of the variations in the major microscopic plasma characteristics: (a) beam acceleration voltage, (b) photomultiplier signal, (c) particle density, (d) x-ray signal, and (e, f) direct and integrated diamagnetic signals.

Measurements of the expansion velocity of the outer boundary of the plasma ''hot electrons'' using miniature thermal probes and x-ray photography of the plasma volume with a pinhole camera showed that the electron-beam plasma ion source retained a sharply defined geometric shape positioned symmetrically about the axis of symmetry of the system and occupied a volume much smaller than that of the vacuum vessel. The radial size of the source in the middle of the magnetic mirror depends mainly on the ratio of τ_c and τ_d. The expansion time is $\tau = \tau_d\tau_c/(\tau_d + \tau_c)$.

The plasma source within which the ''hot electrons'' are concentrated is a volume formed by rotating an ellipse, whose major axis coincides with the axis of the system, about that axis. The axes have lengths of 12 and 30 cm, respectively, and the volume of the hot region was $V = 2800$ cm^3.

At the same time as the geometric size of the source was increasing, its density was rising (oscilloscope trace c) and 400 μs after the start of the beam injection signal, the microwave probe began to be cut off (a probe volume of length 0.8 cm). This indicates an average electron density $n_e \geqslant 1.7 \times 10^{13}$ cm^{-3}. No noticeable attenuation of the 0.4-cm-

wavelength microwave signal was observed under these conditions. Thus, the average electron density in the source was within the range $1.7\times10^{13}\leqslant n_e\leqslant 6.8\times10^{13}$ cm^{-3}.

Trace b of the integrated luminosity of a volume element of the source shows that the luminosity rises rapidly at the beginning and, after reaching a maximum, drops sharply by a factor of two to form a plateau which lasts essentially through the entire beam pulse. The sharp rise in the light signal was associated with the onset of a beam instability and the increased excitation of atoms and molecules prior to breakdown of the gas. The energy radiated by excited atoms and molecules in the breakdown period exceeded the energies radiated in bremsstrahlung and radiative recombination. As the electron energy and charged particle density increased, the number of excited atoms within the source decreased, so that, despite the expansion and enlargement of the optical emission volume, at later times the emission intensity was observed to fall and the main contribution to this emission was from excited atoms and molecules in the boundary layers of the source.

As neutral particles from the walls of the vacuum vessel and the space surrounding the source were incident on the surface of the source, they were ionized over a mean free path λ_0 given by $\lambda_0=V_0/(n\nu_e\sigma_i)$, where $V_0\approx10^6$ cm/s is the component of the neutral particle velocity normal to the emitter axis. Substituting other data from the experiment in this equation yielded $\lambda_0\leqslant0.5$ cm.

One characteristic feature of this source was the presence of a group of electrons with an average energy several times the electron beam energy. The production of these electrons was detected in trace d, which shows the variation in the intensity of hard x rays with energies $E>30$ keV. This emission began when the charged particle density in the source, $n>n_{\text{beam}}$ (roughly 200 μs after injection began), and then the intensity increases monotonically throughout the beam injection time. A study of this emission by standard methods made it possible to determine the temperature T_{er} and density n_{er} of the hot electrons.[11] Thus, spectroscopic studies show that the average energy of the hot electrons was $T_{er}\sim10^5$ eV.

The energy flux Q lost from the discharge through bremsstrahlung was determined using the standard formula $Q=1.7\times10^{-27}nn_{er}T^{1/2}\cdot Z^2$, where n_{er} is the density of electrons with temperature $T_{er}=10^5$ eV, and n is the total charged particle density estimated from the microwave probe data. The measured energy flux was $Q=4.2\times10^{-5}$ J/cm^3. Taking $n\approx2\times10^{13}$ cm^{-3}, we find $n_r\approx4\times10^{11}$ cm^{-3}, i.e., $n_r/n\approx2\times10^{-2}$. The energy of the bulk of the electrons in the discharge measured by the multigrid probe was 100 eV.

Oscilloscope traces e and f show the variation in the unintegrated and integrated diamagnetic signals and determine the rise in the gas kinetic pressure in the source. The measured gas kinetic pressure in the source is $W=3\times10^{16}$ eV/cm^3. It is clear from these traces that heating occurs essentially throughout the electron beam pulse. The energy lifetime of the source was $\tau_c=1.5\mu$s and the energy in the source began to decrease when the beam acceleration pulse began to decay. The energy lifetime is associated with the high energy electrons, which have a small cross section

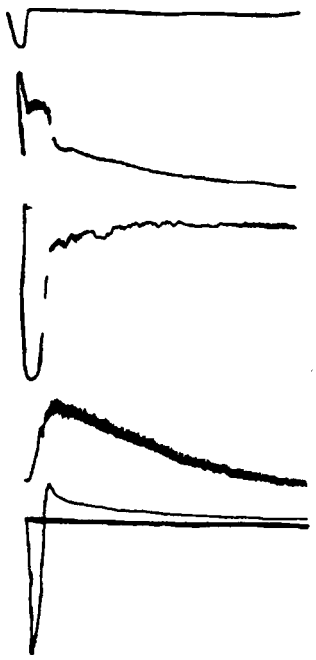

FIG. 2. Oscilloscope traces of the variations in the macroscopic plasma characteristics during the heating and cooling phases. The scan duration is 0.1 s.

for scattering on the ions, and, therefore, a long collision time $\tau_{ei}=AT_{er}^{3/2}/n$, where A is a constant. They could, therefore, be confined for a long time $\tau\geqslant10^{-3}$ s by the magnetic system and contained a large fraction of the thermal energy of the source.

Traces d and f show that the charged particle lifetime τ_e is much greater than the time of flight τ_0 of the neutral particles across the system, so the charged particle density in the system could become greater than the initial neutral density, i.e., plasma buildup was observed. After the electron beam energy was turned off, the charged particles in the source are left to themselves and begin to leave the system. The series of oscilloscope traces in Fig. 2 illustrates the variation of the same macroscopic plasma characteristics which have been analyzed in the active phase of the discharge.

The decay characteristics of the source imply that the rates of fall of the luminosity and charged particle density are substantially higher than the decay times for the x-ray bremsstrahlung and source energy, since the cold electrons in the source have large cross sections for interactions with the neutral gas and source ions and, therefore, shorter times between collisions, i.e.,

$$\frac{\tau_{e\Gamma}}{\tau_{ex}}=\left(\frac{T_{e\Gamma}}{T_{ex}}\right)^{3/2}\gg1.$$

In each collision event, the velocity vector of a charged particle changed by an amount comparable to magnitude of the velocity itself, i.e., the charged particle entered the loss cone ($\alpha=\arcsin\sqrt{B_0B}$, where B_0 and B are the magnetic fields in the center of the mirror and in the center of the solenoid, respectively; $\alpha=35°$ for the magnetic system (mir-

ror)) and the particle left the mirror volume, since the mean free path of the particle exceeded the geometric dimensions of the mirror.

The oscilloscope traces of the x-ray emission illustrate the long confinement of the hot electrons resulting from their small cross sections for scattering by the ions and neutral gas. The time variation in the diamagnetic signal is similar to that in the x-ray probe signal.

All of the above confirms once again the conclusion that the bulk of the thermal energy of the source is contained in a group of ''fast'' electrons.

The deep layers of the source are protected from the flux of neutral particles and the neutrals are multiply ionized by the bulk of ''cold'' electrons in the source.

Therefore, the electron beam transfers its energy to ionization and heating of the working gas in the source with a high efficiency ($\sim$60 %). As the source interacted with the beam, it remained macroscopically stable and the energy lifetime in the active phase was 10^{-3} s, and in the passive phase, 10^{-1} s. The mean free path of the neutral particles is considerably shorter than the source dimensions, so ionization took place mainly in the boundary regions and charged particles were observed to accumulate in the system.

This source has been used for shaping and accelerating multiampere ion beams.

[1] S. M. Blokhin and V. V. Kolesnikov, Zh. Tekh. Fiz. **65**(5), 183 (1995) [Tech. Phys. **40**, 507 (1995)].
[2] A. V. Voĭtkiv and A. V. Koval', Zh. Tekh. Fiz. **65**(11), 12 (1995) [Tech. Phys. **40**, 1097 (1995)].
[3] I. G. Ivanov, E. L. Latush, and M. F. Sém, *Metal Vapor Ion Lasers* [in Russian], Énergoatomizdat, Moscow (1990), 255 pp.
[4] Yu. Ya. Valkolupov, *Materials from the 11th All-Union Conference on Coherent and Nonlinear Physics* [in Russian], Izd. Erevanskogo Universiteta, Erevan (1982), pp. 16–17.
[5] L. L. Zakatov, A. G. Plakhov, D. D. Ryutov, and V. V. Shapkin, Zh. Éksp. Teor. Fiz. **54**, 1098 (1976) [*sic*].
[6] R. A. Demirkhanov, A. G. Gevorkov, A. F. Popov, and O. A. Kolmakov, Zh. Tekh. Fiz. **45**, 1862 (1975) [Sov. Phys. Tech. Phys. **20**, 1179 (1975)].
[7] Ya. B. Faĭnberg, A. K. Berezin, and T. P. Berezina, Atomnaya Énergiya **11**, 403 (1961).
[8] A. N. Karkhov, Zh. Éksp. Teor. Fiz. **56**, 792 (1969) [Sov. Phys. JETP **29**, 431 (1969)].
[9] É. I. Kuznetsov and D. A. Shcheglov, *High-Temperature Plasma Diagnostic Techniques* [in Russian], Atomizdat, Moscow (1974), 152 pp.
[10] N. G. Baranov, Yu. Ya. Volkolupov, and A. D. Kuribeda, Radiotekhnika, No. 39, pp. 15–28 (1976).
[11] Yu. N. Luk'yanov, *Hot Plasmas and Controlled Thermonuclear Fusion* [in Russian], Nauka, Moscow (1975), 407 pp.

Translated by D. H. McNeill

Modeling the development of the stepped leader of a lightning discharge

A. A. Dul'zon, V. V. Lopatin, M. D. Noskov, and O. I. Pleshkov

Scientific Research Institute of High Voltages, Tomsk Polytechnical University, 634050 Tomsk, Russia
(Submitted December 10, 1997; resubmitted March 13, 1998)
Zh. Tekh. Fiz. **69**, 48–53 (April 1999)

A stochastic–deterministic model is presented for the propagation of a downward-moving leader. Lightning formation is described by a stochastic growth of branching discharge channels which is determined by the electrostatic field. The dynamics of the electric field and of the charge distribution over the lightning structure are calculated deterministically. The model includes the initiation of lightning, a preliminary discharge in a cloud, the propagation of a downward-moving stepped leader toward the earth, and the initiation and upward motion of a return stroke from the earth's surface. Numerical execution of the model yields a dynamic picture of the development of the downward-moving leader and of the intracloud discharge structure. The effect of the charge density in the cloud and the parameters of the developing channels on the spatial–temporal, current, and charge characteristics of the stepped leader's propagation are studied. The effect of free-standing structures on the distribution of points on the earth's surface where lightning strikes is examined. © *1999 American Institute of Physics.*
[S1063-7842(99)00904-6]

INTRODUCTION

Conducting cloud–earth lightning channels develop as a result of the stochastic propagation of a stepped leader toward the earth. The major factor responsible for the dynamics of this process is the electric field strength near the leader tip.[1–3] During the first stages of propagation of the leader, the field strength at its tip depends on the charge in the leader channel and the volume charge distribution in the cloud. As the leader moves toward the earth, the buildup of charge on the earth's surface and on objects protruding above it begins to have an ever greater effect on the field distribution. As the leader approaches a critical distance (the striking distance) from the earth, a return leader may develop. Mathematical models of the propagation of the downward-moving stepped leader including the parameters of the thundercloud, atmosphere, and underlying surface is not just of scientific interest, but also is of practical value for the improving lightning protection systems.

There are currently several different models for the propagation of the stepped leader.[4–14] Numerical implementations of these models have been used to study the distribution of lightning strike points over the earth's surface,[4,9,12] determining the probability of damaging various earthbound and flying objects,[7,8,10–12] and calculating the protection zones for lightning arrestors.[7,8,10,12]

However, the existing models employ major simplifications. They describe the motion of the stepped leader only between the based of a cloud and the earth, without considering the development of intracloud lightning channels, the properties of the discharge channels are assumed to be time independent, and the redistribution of charges in the cloud and the channels as the leader moves is neglected.

In this paper we propose a stochastic–deterministic model for the propagation of a stepped leader based on a modification of a fractal model for dielectric breakdown.[14,15] The formation of the intracloud structure of the lightning and the earthward propagation of the stepped leader are described by a stochastic growth in the fractal structure controlled by the electric field strength. The dynamics of the distributions of the electric field and charge are determined deterministically on the basis of Maxwell's equations and suitable semiempirical formulas for the conductivity of the lightning channels.[16] Unlike the existing fractal models for the development of the stepped leader,[6,12] which are based on the law for fractal structure growth proposed by Niemeyer *et al.*,[14] in our model currents flowing along the discharge channels are introduced.[17] A numerical execution of this model is used to study the spatial–temporal, current, and charge characteristics of the stepped leader's development, as well as the distribution of strike points over the earth's surface.

MODEL FOR DEVELOPMENT OF A STEPPED LEADER

The model describes the development of a negatively charged stepped leader moving from a cloud to the earth and the development of a return leader from the earth. Thus, the model region includes only the lower, negatively charged region of the cloud, the part of the earth's surface located below the cloud, and the space between them (Fig. 1). The initiation site for the lightning is assumed to lie between the center of the negative charge distribution and the region of localized positive charge in the lower part of the cloud. This assumption is justified by the possible existence of local positively charged regions in the lower part of clouds[1–3] and by data on the location of the rf signals accompanying the initiation of a lightning discharge.[18–20] The development of the discharge channels within the cloud and in the earthward direction is described by stochastic growth steps. The prob-

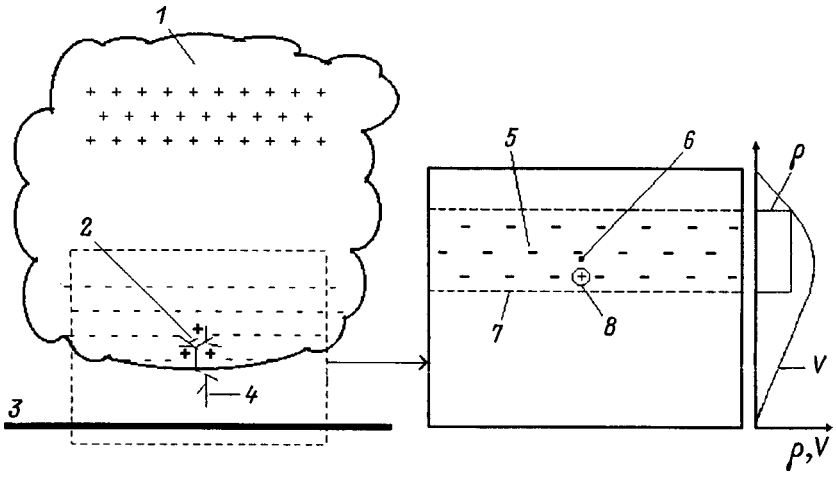

FIG. 1. The model geometry: (*1*) thundercloud, (*2*) intracloud discharge, (*3*) earth's surface, (*4*) stepped leader, (*5*) negatively charge region, (*6*) lightning initiation point, (*7*) cloud base, (*8*) positive charge; V and ρ are the vertical distributions of the potential and negative charge density.

ability P of a step's growing in some direction **n** is determined by the state of the medium and the projection of the local electric field E_n in this direction,

$$P=\begin{cases}\dfrac{(E_n)^{\eta}}{Z}, & E_n \geqslant E_c,\\ 0, & E_n < E_c,\end{cases} \tag{1}$$

where $Z=\Sigma(E_n)^{\eta}$ is a normalizing factor. (The sum is taken over all possible growth directions).

The critical field E_c and growth parameter η depend on the propagation conditions. The critical field E_c in the cloud is assumed to be 5 times smaller than E_c in air because of the substantial reduction in the breakdown voltage in air when ice crystallites and water vapor and droplets are present, as well as because of the pressure drop with altitude. The growth parameter η can be estimated by analyzing the leader propagation mechanism. It is logical to related the growth probability to the intensity of energy release in the streamer zone. In that case, the growth parameter should be equal to two ($\eta=2$).

At each growth step the electric field distribution is found using Gauss' theorem

$$\operatorname{div}(\varepsilon\varepsilon_0\mathbf{E})=\rho, \tag{2}$$

where ε is the dielectric constant, ε_0 is the permittivity of free space, and ρ is the density of the free charge distribution in the lightning channels and cloud.

As first approximation in the model, we assume a uniform initial distribution of negative charge in the lower part of the cloud (Fig. 1). This approximation differs substantially from the classical dipole model for the electrical structure of a cloud and corresponds more to the field distribution created by the charges in several thunderstorm cells.

The boundary conditions for solving Eq. (2) are the fixed potentials on the upper and lower boundaries of the model region and the free potentials at the side boundaries (Fig. 1). The potential of the lower boundary corresponding to the earth's surface is assumed to be zero. The upper boundary lies between the positively and negatively charged parts of the thundercloud. Its location also corresponds to a zero potential. The potential of the side boundaries is recalculated in each growth step using the condition that the flux of the field strength vector at them is zero.

The charge distribution varies with time. Charge is collected in the storm cloud and charge is carried to the earth by the propagating stepped leader. The characteristic time from lightning initiation until the leader reaches the earth is approximately 0.02 s.[1–3,21] Since this time is much shorter than the Maxwell relaxation time of the charges in the cloud and in the air, the conduction currents in the latter can be neglected. Charge is collected in the cloud only by the growing intracloud lightning structure.

The charge distribution obeys the continuity equation

$$\frac{d\rho}{dt}=-\operatorname{div}\mathbf{j}, \tag{3}$$

where **j** is the current density in the lightning channels.

Charge transfer along the lightning channels obeys Ohm's law

$$I=\gamma\cdot E_l. \tag{4}$$

Here I is the current, γ is the conductivity of a unit channel length (the running conductivity), E_l is the drop in the field per unit channel length. γ is equals the product of the specific conductivity of the plasma in the leader channel, σ, and the effective transverse channel cross section S ($\gamma=\sigma\cdot S$). The variation in the conductivity γ is determined by the balance between the production and dissipation of Joule heat in the lightning channel.[3,21] As a first approximation it is assumed that the increase in the conductivity is directly proportional to the energy release, i.e.,

$$\frac{d\gamma}{dt}=\xi\gamma\cdot E_l^2, \tag{5}$$

where ξ is the growth rate parameter for the conductivity.

Equation (5) is a differential analogy of the Rompe–Weitzel formula for the conductivity of a long spark in gases. The conductivity per unit length of a newly formed channel is taken to be γ_0. For an adequate description of the dynamics of the charges and of the change in the conductivity in the model, we introduce the physical time t. Each stage in the

growth of the lightning structure corresponds to an interval of physical time Δt. It follows from Eq. (1) that the probability ω of channel growth per unit time has a power law dependence on the field strength, i.e.,

$$\omega(E)=\frac{E^{\eta}}{\Theta}, \tag{6}$$

where Θ is a time constant.

Assuming that the magnitudes of the growth in the different channels are statistically independent of one another, we find that the probability W that any one of the lightning channels will grow during the time unit Δt is determined by summing over all the possible directions $W=\Sigma\omega(E)$. The time interval corresponding to a growth step is inversely proportional to W, so that

$$\Delta t=(W(E))^{-1}=\frac{\Theta}{\Sigma E^{\eta}}=\frac{\Theta}{Z}. \tag{7}$$

Therefore, the time constant Θ has the significance of a transition parameter to physical time.

The propagation of the stepped leader causes the field strength near the earth's surface to increase. When the field strength reaches the critical level, a return leader is initiated from the surface. The propagation of the return leader obeys the same probability dependence (1) on the field strength as the stepped leader and they cease to develop further after they meet. The strike point of the lightning is assumed to be the place where the return leader is initiated.

In the numerical simulations the model was executed numerically on a two dimensional square mesh. The size of the mesh corresponds to the average length of the leader steps. The lightning structure develops on one edge or the diagonal of the mesh in a single step in accordance with the probability distribution (1). The distributions of the electric fields and charges are calculated using a finite-difference approximation to Eqs. (2)–(7).

RESULTS AND DISCUSSION

The model proposed here has been used to study the spatial–temporal and charge characteristics of a developing leader in a lightning discharge. The ranges of variation of the model parameters were chosen on the basis of experimental data from the literature and a comparison of the morphogenesis of the calculated patterns with observations in nature. The computer model can therefore be regarded as a method for solving the inverse problem that yields additional parameters beyond the natural observations. A two-dimensional calculation yields dimensionless quantities which can only be compared qualitatively with measurement data.

A stepped leader is initiated as a result of a prior breakdown between a negatively charged region of a cloud and a local positive charge located near the cloud base or fluctuation in the negative charge (Fig. 2a). At first the preliminary breakdown channel propagates vertically, but then side branches begin to develop. The resulting pattern is qualitatively consistent with the data of VHF lightning detection and ranging (LDAR) as to the location of the initial lightning development stage;[18,19] this is evidence of the validity of the model. The intracloud discharge structure collects charge efficiently and raises the field strength in its lower part. When the field strength exceeds the critical field strength E_c (air), it becomes possible for an earthward stepped leader to propagate. A delay in the development of the stepped leader relative to the onset of the intracloud discharge has been noticed by many observers.[19,20] The development of the intracloud structure continues predominantly in a horizontal direction, and the rate of channel branching in it is higher than for the channels of the stepped leader. The propagation of the stepped leader from the cloud base to the earth is also accompanied by random branching (Fig. 2b). As a rule, one of the branches of the leader begins to predominate and overtakes the others (Fig. 2c). The approach of the tip of the dominant branch to the earth causes an rise in the field strength at the underlying surface. As a result of this, a return leader sets out from the surface (Fig. 2d). If tall structures exist at the surface, then they will be places where a return leader starts. In this case, initiation occurs sooner and the return leader travels a larger distance before it meets the stepped leader (Fig. 3). The calculated patterns are consistent with generally accepted ideas and observations. The distribution of lightning strike points over a plane surface is close to normal. The peak of the distribution lies under the lightning initiation point (Fig. 4a). The existence of tall structures on the surface causes a significant redistribution of the strike

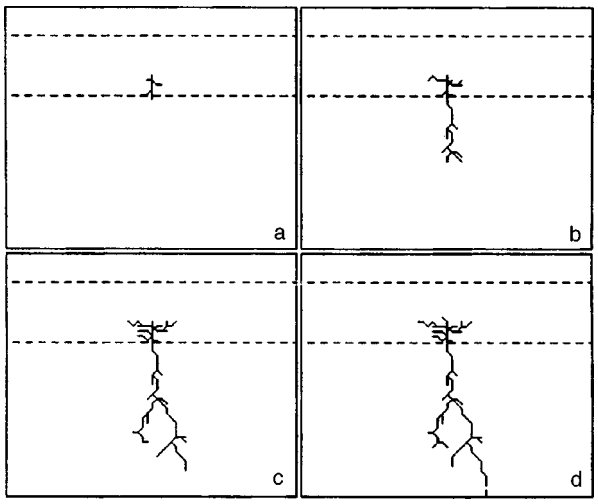

FIG. 2. The development of a stepped leader and of the intracloud lightning structure. Model parameters: $\rho=0.6$, $\gamma_0=20$, $\xi=1$, $\Theta=5$, $E_c(\text{air})=2.5$, and $E_c(\text{cloud})=0.5$.

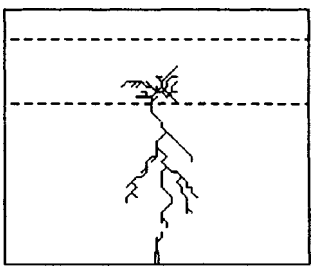

FIG. 3. The development of a return leader from a tall structure with $h=2$. The model parameters are the same as in Fig. 2.

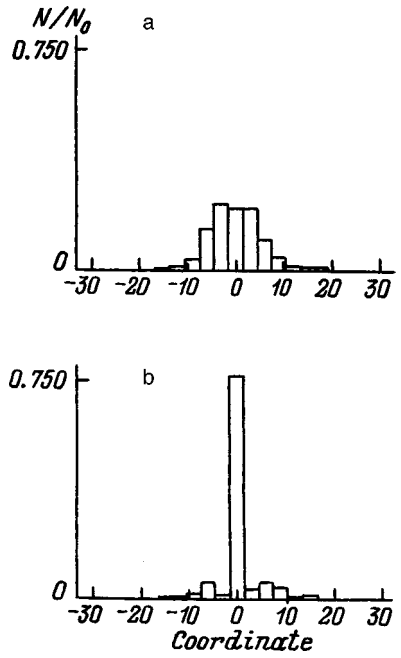

FIG. 4. Distributions of lightning strike points over a surface. The number of trials is $N_0=300$. The model parameters are the same as in Fig. 2. (a) Plane surface, (b) with a tall structure ($h=2$) located under the lightning initiation point.

points (Fig. 4b). Then the probability of damage to a structure becomes greater as its height increases and when it lies closer to the position below which the lightning was initiated (Fig. 5).

The development of the stepped leader is accompanied by polarization of the entire lightning discharge structure. Its intracloud portion is charged positively, while the part moving toward the ground is negatively charged. The linear charge density of the leader increases exponentially on approaching its tip. (There are no unambiguous published data on the charge distribution along the channel of a stepped leader. A uniform charge density distribution and one that decreases exponentially with altitude have been proposed.[3]) The redistribution of the charges over the lightning structure

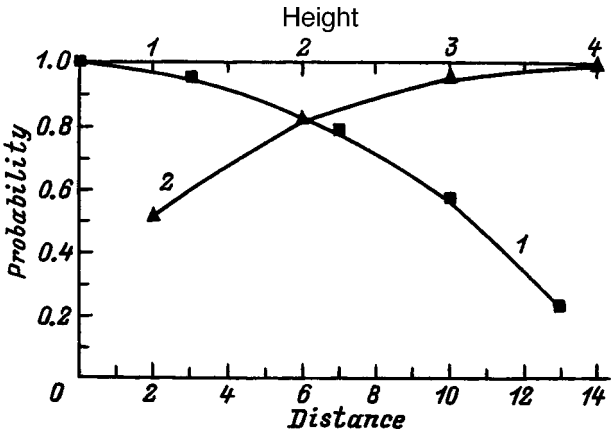

FIG. 5. The probability of damage to a tall structure as a function of its position relative to the coordinates of the initiation point (*1*) and of its height above the center (zero distance) (*2*). The points correspond to data averaged over 200 computer runs.

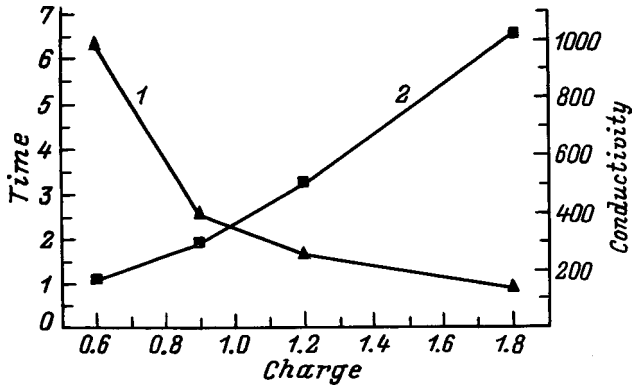

FIG. 6. The time for a stepped leader to develop (*1*) and its maximum running conductivity (*2*) as functions of the initial charge density in a cloud. $\gamma_0=20$, $\xi=1$, $\Theta=5$, $E_c=2.5$ (air) and 0.5 (cloud). The points correspond to data averaged over 200 computer runs.

causes part of the cloud potential to be carried away to the tip of the stepped leader and raises the potential of the cloud near the site of the intracloud lightning channels. The field strength at the leader tip increases, while it decreases in the channels. After the leader has covered half the distance between the cloud and the ground, the field at the tip of the dominant branch is roughly an order of magnitude higher than the average field strength between the cloud and the ground, while the field strength in the leader channels is an order magnitude lower than the average field strength. These relationships among the fields are consistent with existing ideas regarding the distribution of the fields in the stepped leader channel.[3,20] The highest current and conductivity are attained in the channels of the dominant branch of the stepped leader. The currents increase as the leader approaches the earth. The channels of the intracloud lightning structure have a lower conductivity than the stepped leader channels. Efficient transport of charge to the leader channel initiation point ensures a ''thick'' intracloud discharge structure.

The characteristics of a stepped leader depend to a great extent on the cloud parameters and the features of the environment in which the lightning discharge takes place. In terms of our approach, by varying the model parameters it is possible to reproduce the spreads in the currents and velocities of downward propagating leaders observed in nature. These simulations show that the magnitude of the charge in the cloud has the greatest effect on the spatial–temporal and current characteristics of the stepped leader's development. Increasing the charge on the cloud leads to a proportional increase in the average field strength between the cloud and the earth. As a result of the drop in the time for the stepped leader to develop, the maximum running conductivity of the channel (Fig. 6) increases, while the discharge structure becomes more strongly branched. Increasing the conductivity growth parameter ξ from 0.2 to 5 yields similar consequences, but this effect is not as strong qualitatively, and changing the initial channel conductivity γ_0 by an order of magnitude essentially had no effect on the development of the leader.

The distribution of the time interval from the time the

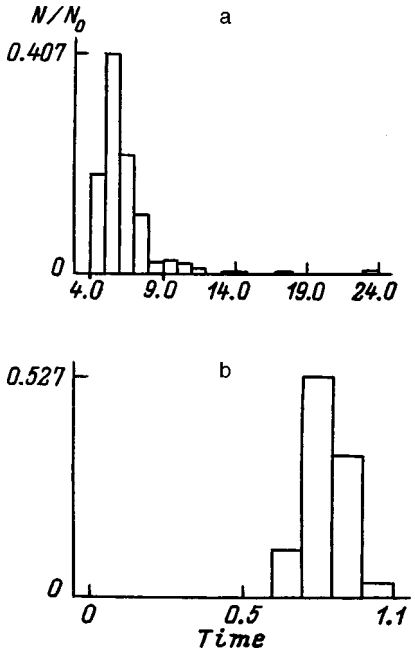

FIG. 7. Some distributions of the time for a stepped leader to develop for different model parameters. The number of runs $N_0=300$; $\rho=0.6$ (a), 1.8 (b); the remaining parameters are as in Fig. 6.

lightning is initiated to the time the leader reaches the ground is shown in Fig. 7 for different model parameters. The propagation velocities of the stepped leader differ by roughly an order of magnitude in these cases. Based on the relationship between the velocities and the character of branching in them, slow and fast leaders can be compared with observed α- and β-stepped leaders.[2,21] The leader propagation velocity increases slightly on approaching the ground. Figure 8 shows typical plots of the length of the dominant lightning branch as a function of time for the two types of leader. The current in the fast leader is roughly an order of magnitude higher than in the slow leader. The experimentally observed spread in the current in stepped leaders is also roughly an order of magnitude.[1,3]

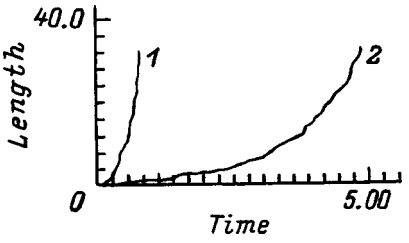

FIG. 8. The length of the dominant branch of the leader as a function of time. $\rho=1.8$ (*1*), 0.6 (*2*); the remaining parameters are as in Fig. 6.

Our model, therefore, provides an adequate description of the basic features of the development of a stepped leader. A simulation was in qualitative agreement with experimental data. The model can serve as a basis for creating new, efficient computational techniques for calculations of the protection zone of lightning arrestors and of the probability of damage to various objects as functions of their geometry, the parameters of the underlying surface, and the characteristics of a thundercloud.

[1] E. P. Krider, in *The Earth's Electrical Environment*, National Academy Press, Washington (1986), p. 263.
[2] A. M. Uman, *The Lightning discharge*, Academic Press, London (1987), p. 397.
[3] E. M. Bazelyan and Yu. P. Raizer, *Spark Discharges* [in Russian], Izd. MFTI (1997), 320 pp.
[4] L. Niemeyer, J. Phys. D **20**, 897 (1987).
[5] Z. Kawasaki, M Nakano, T. Takeuti *et al.*, in *Proceedings of the International Conference on Atmospheric Electricity* (1988), pp. 483–488.
[6] Z. Kawasaki, K. Matsuura, M. Nakano *et al.*, Res. Lett. Atmospheric Electricity **9**, 63 (1989).
[7] L. Dellera and E. Garbagnati, IEEE Trans. Power Deliv. **PD-5**, 2009 (1990).
[8] L. Dellera and E. Garbagnati, IEEE Trans. Power Deliv. **PD-5**, 2023 (1990).
[9] T. Takeuti, T. Hashimoto, and N. Takagi, J. Atmos. Electr. **13**, 9 (1993).
[10] T. Takeuti, T. Hashimoto, and N. Takagi, in *Eighth International Symposium on High Voltage Engineering*, Yokohama, Japan (1993), pp. 265–267.
[11] A. S. Gaixorovsky and K. V. Karasyuk, in *Eighth International Symposium on High Voltage Engineering*, Yokohama, Japan (1993), pp. 277–280.
[12] N. I. Petrov and G. N. Petrova, Pis'ma Zh. Tekh. Fiz. **18**(3), 14 (1992) [Sov. Tech. Phys. Lett. **18**, 65 (1992)].
[13] L. Ruhnke and V Mazuk, in *Tenth International Conference on Atmospheric Electricity*, Osaka, Japan (1996), pp. 192–195.
[14] L. Niemeyer, L. Pietronero, and H. J. Wiesmann, Phys. Rev. Lett. **52**, 1033 (1984).
[15] V. R. Kukhta, V. V. Lopatin, and M. D. Noskov, Zh. Tekh. Fiz. **65**(2), 63 (1995) [Tech. Phys. **40**, 150 (1995)].
[16] A. Dulson, M. Noskov, V. Lopatin and D. Shelukhin, in *Tenth International Conference on Atmospheric Electricity*, Osaka, Japan (1996), pp. 260–263.
[17] V. Lopatin, M. Noskov, and O. I. Pleshkov, in *Proceedings of the 12th International Conference on Gas Discharges and Their Application*, Greifswald, Germany (1997), pp. 432–435.
[18] P. L. Rustan, M. A. Uman, D. G. Childer *et al.*, J. Geophys. Res. **85**, 4893 (1980).
[19] D. E. Proctor, J. Geophys. Res. **88**, 5421 (1983).
[20] L. Maier, C. Lennon, P. Krehbiel *et al.*, *Tenth International Conference on Atmospheric Electricity*, Osaka, Japan (1996), pp. 280–283.
[21] E. M. Bazelyan, B. N. Gorin, and V. I. Levitov, *Physical and Engineering Foundations of Lightning Protection* [in Russian], Gidrometeoizdat, Leningrad (1978), p. 222.

Translated by D. H. McNeill

Determination of the recombination parameters of a semiconducting material by the Prony method

Yu. A. Bykovskiĭ, K. V. Kolosov, V. V. Zuev, A. D. Kiryukhin, and S. I. Rasmagin

Moscow State Engineering Physics Institute (Technical University), 115409 Moscow, Russia
(Submitted October 13, 1997)
Zh. Tekh. Fiz. **69**, 54–59 (April 1999)

A generalized Prony method is used to process experimental data of the relaxation type. The method can be used to separate a compound relaxation process into elementary components (damped exponentials) and to systematically trace the variation of their parameters in the presence of external excitation, thereby opening additional channels of information on the generating physical process. The method is applied to the processing of photoconductivity decay data for complex-doped, compensated silicon in the temperature range 100–300 K after pulsed excitation by laser radiation with $\lambda=1.06$ μm in a microwave field. The position of the energy level is obtained as 0.17 ± 0.01 eV, in good agreement with the energy of the oxygen-vacancy complex. © *1999 American Institute of Physics.* [S1063-7842(99)01004-1]

INTRODUCTION

The level of development of the production technology of semiconductor devices in today's world is a major factor governing the development of industry at large. There is not a single branch of the national economy in which various semiconductor elements, sensors, and microprocessors would not be used. Devices are manufactured on the basis of high-quality semiconducting materials with predetermined properties that remain stable in the face of selectively targeted external influences.

These demands raise the issue of reliable, on-line methods for determining various parameters of a semiconducting material prior to the fabrication of structures and instruments in which it is to be used. Contactless measurement techniques based on the exploitation of special characteristics of the interaction of electromagnetic radiation with matter have gained widespread acceptance at the present time.[1–5] The interaction of a microwave field with a semiconductor is utilized on an especially broad scale for the contactless determination of the parameters of a semiconducting material.[3–5]

In this paper we discuss aspects of determining the recombination characteristics of a semiconducting material (in the example of complex-doped silicon) by pulsed laser irradiation in a microwave field. The relaxation curves are processed by the Prony method (a method for the simultaneous estimation of exponential parameters[6]).

It is important to note that the burgeoning development of computer engineering has now made it possible for highly efficient computer algorithms to be used in real time on comparatively low-cost general-purpose computers.[7,8]

BASIC PHENOMENOLOGICAL MODELS

The interaction of a microwave field with a conducting material is known to be governed largely by the conductivity in the bands of the material.[9] Conditions can therefore be established whereby the power of microwave reflections from a semiconductor is characterized by the density of both equilibrium and nonequilibrium carriers in the allowed bands of the semiconductor.

Existing phenomenological models can be used to relate the characteristic decay time of the nonequilibrium carrier density to impurity parameters: the position of the impurity energy level, the recombination coefficient, and the impurity concentration.[10] Levels in the band gap can be replaced by carriers both from an allowed band (attachment level) and from two bands (recombination level). If the investigated level is a recombination level and if the density N_a of a compensating acceptor impurity in *n*-type silicon is greater than the density of shallow donors, the temperature dependence of the time constants τ_n and τ_p is dictated by the dependence of the recombination coefficients for electrons γ_n and holes γ_p:

$$\tau_n=\frac{1}{\gamma_n(N_a-N_d)}; \quad \tau_p=\frac{1}{\gamma_p N_d}. \tag{1}$$

In the case of α-attachment, i.e., when the equilibrium of a level with a band is established in a time shorter than the carrier lifetime, the temperature dependence of the nonequilibrium carrier lifetime τ can have the form[10]

$$\tau=\tau_p\left(1+\frac{M}{N_{CM}}\right), \quad N_{CM}=N_c\cdot\exp\left(-\frac{E_M}{kT}\right), \tag{2}$$

where τ_p is the recombination time through a deep level of the type (1), N_{CM} is a characteristic density in the band if the Fermi level is situated at an attachment level, N_C is the density of states in the conduction band, M is the density of attachment levels, and E_M is the energy position of the attachment level.

Consequently, the attachment energy level for $(M/N_{CM})\gg 1$ can be determined from the temperature dependence of τ. It is important to note that if more than one

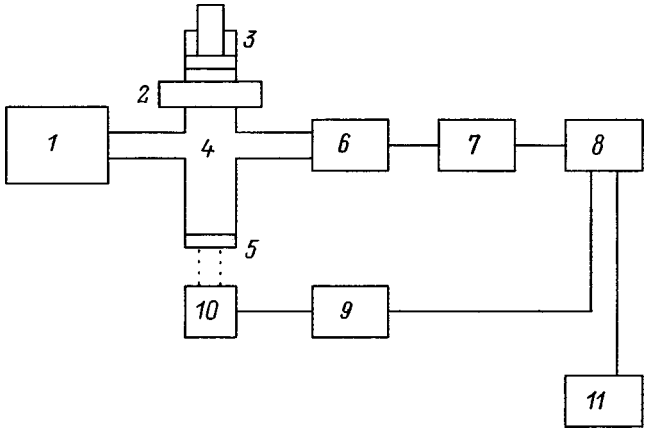

FIG. 1. Block diagram of the experimental arrangement. (*1*) microwave oscillator; (*2*) attenuator; (*3*) short-circuited piston; (*4*) T-bridge; (*5*) sample in the cryostat; (*6*) microwave diode; (*7*) amplifier; (*8*) ADC; (*9*) synchronizing unit; (*10*) laser with light-emitting diode; (*11*) computer.

impurity level appears at a given temperature, the photoconductivity decay curve can be expressed by a linear combination of exponential terms:

$$f(t)=\sum_{i=1}^{N} a_i \exp\left(-\frac{t}{\tau_i}\right), \tag{3}$$

where a_i denotes the partial amplitudes, and τ_i are the partial relaxation times.

EXPERIMENTAL CONDITIONS

Gold-doped samples of as-grown KÉF-3.5 silicon were prepared for the experiment. The relaxation time of the photoconductivity in the as-grown sample was 130 μs. Gold was deposited on the substrate by vacuum evaporation, and thermal diffusion was performed in a vacuum furnace at 900 °C. After heat treatment and preliminary inspection the sample was placed in the experimental apparatus shown in Fig. 1.

Short (<80 ns) laser pulses with $\lambda=1.06$ μm were directed onto the sample *5* in a cryostat ($T=100-300$ K), generating nonequilibrium carriers in the sample. Since the nonequilibrium conductivity of the sample varied with time, the intensity of the reflected microwave ($\lambda=3$ cm, $P<2$ mW) was modulated in accordance with the decay time of the photoconductivity. The parameters of the microwave field and the laser beam were chosen so that the microwave signal of the diode *6* would be proportional to the conductivity of the sample. The signal was then transmitted through the amplifier *7* and the eight-bit amplitude-to-digital converter (ADC) *8* and was finally recorded as discrete readings in the computer *11*.

Inasmuch as the characteristic decay times of the photoconductivity of the doped sample were $0.1-30$ μs, the parameters of the ADC were chosen so as to achieve unambiguous recording of the entire frequency range of the signal[6,11] (the discretization step was 50 ns, and the number of readings on the decay curve was 2048). The relaxation curves were recorded at various temperatures and were then analyzed to assess the feasibility of determining the impurity parameters of the dopant and entrained secondary dopant elements.

PRONY METHOD

The problem of determining the parameters of Eq. (3) can be solved formally by the least-squares method, provided only that the number of readings in the data sample exceeds the number of parameters estimated. In reality, however, a nonlinear system of transcendental equations is encountered in solving this kind of problem, even in the simplest case, and it is problematical whether the system is even solvable by numerical methods (to say nothing of the colossal expenditure of computational resources that would be required).[6,7] The Prony method can be used to reduce the problem to two linear least-squares procedures, the first of which determines the relaxation parameters, while the second determines the partial amplitudes. A detailed description of the Prony method can be found in the relevant literature (see, e.g., Ref. 6). We merely call attention to the number of terms in Eq. (3), because it is chosen *a priori*.

The values of certain parameters τ calculated by the Prony method can be complex. This means that the raw data contain oscillatory components. Terms of this kind are generated by the inevitable noise present in the experimental data. Ostensibly, therefore, the larger the terms in Eq. (3), the more accurately the Prony method can approximate the experimental noise and the more accurate will be the values obtained for the components of the useful signal. Unfortunately, this statement is true up to a limit, because the Prony method becomes unstable for a large number of parameters by virtue of the fact that the systems of linear equations whose solution is the very basis of the method become poorly conditioned.[6]

Our investigation has been carried out for ten terms. We have arrived at this particular choice, first of all, by drawing on cumulative experience and, second, by taking into account *a priori* information about the investigated phenomenon.

SIMULATION RESULTS

To draw some conclusion as to the applicability of the Prony method for processing our experimental data, it has been necessary to run a preliminary simulation.

For this purpose a sequence of 20 samples of relaxation curves was generated artificially, each curve having the form

$$f(t_i)=a_f \exp\left(-\frac{t_i}{\tau_f}\right)+a_s \exp\left(-\frac{t_i}{\tau_s}\right), \tag{4}$$

where τ_f and τ_s are the partial relaxation times (fast and slow times, respectively), a_f and a_s are the corresponding partial amplitudes, $t_i=\delta\cdot i$ is the reading time, $\delta=50$ ns, $i=0\ldots2047$.

The functional relation (4) symbolizes the experimental curve, so that the parameters τ_f, τ_s, and δ have the dimensions of time. In turn, τ_f, τ_s, a_f, and a_s depend on the

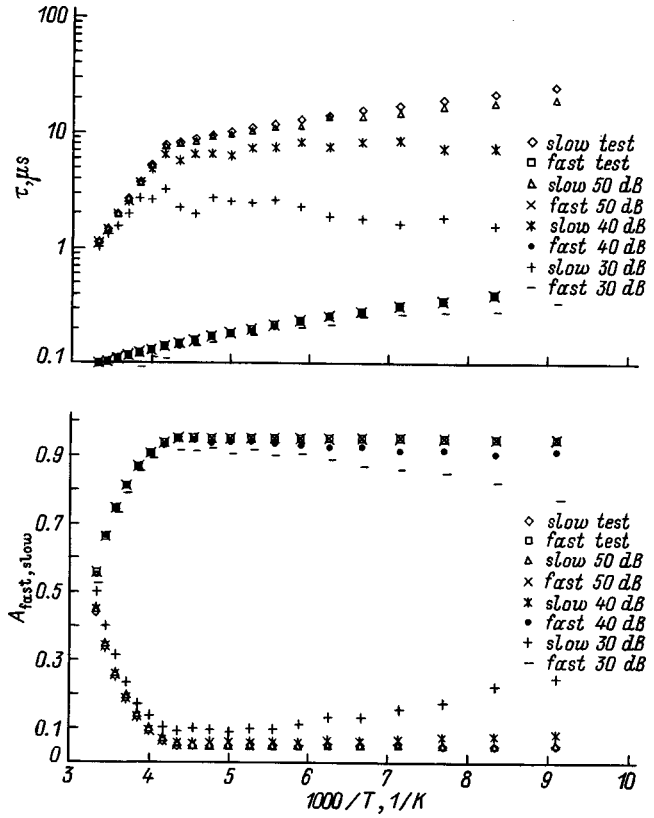

FIG. 2. Simulation results (data sample + white noise). Comparison of the results by the Prony method with a test set of parameters. The signal-to-noise ratio varies from 30 dB to 50 dB.

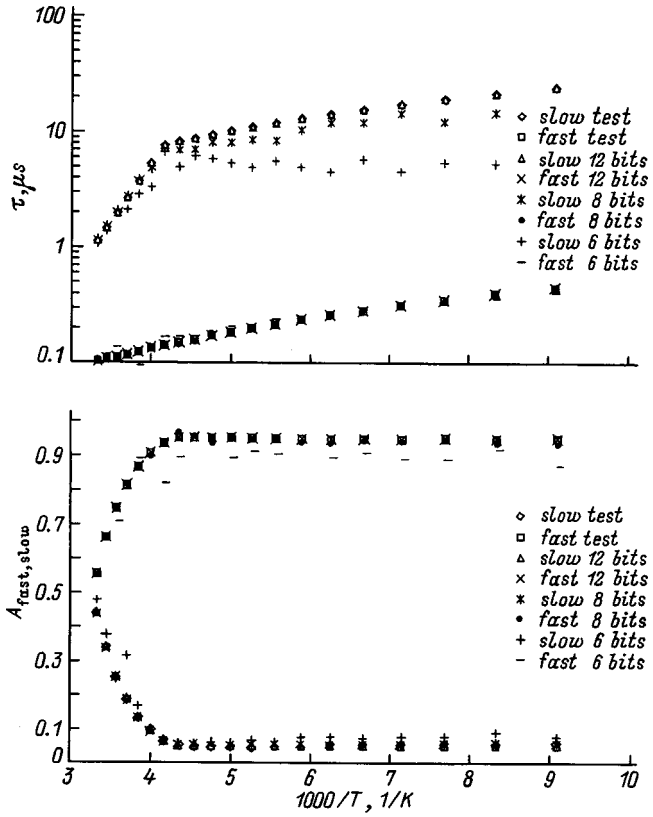

FIG. 3. Simulation results (data sample + discretization noise). Comparison of the results by the Prony method with a test set of parameters. The analysis covers 6-, 8-, and 12-bit ADCs.

parameter T (which denotes the temperature); the corresponding dependences are shown in Figs. 2 and 3 (slow test and fast test).

A stochastic process (white noise with different variances σ) was initially added to the deterministic process (4) of each raw data sample. The resulting set of samples was thus processed by the Prony method. The processing shows that the parameters determined by the Prony method agree with the test set within 0.1–3% error limits except for the parts where the partial amplitude of some exponential function become smaller than the fivefold (14 dB) variance σ of the white noise (Fig. 2, slow test). Here the total signal amplitude must be at least 100 times σ (40 dB greater).

We have also simulated the application of analog-to-digital conversion to the sequence of raw data samples, for various ADC word lengths. Here we assumed that the total signal amplitude corresponds to half the word length. The simulation results are shown in Fig. 3. It is evident that the Prony parameters agree with the initial parameters within 3% error limits (except for the small-amplitude parts) when an ADC word length of at least eight bits is used.

The simulation has thus shown that the Prony method yields satisfactory results (i.e., the errors of the method do not exceed the instrumental error) in the solution of problems with four parameters if at least an eight-bit analog-to-digital converter is used and if the total amplitude-to-noise ratio is at least 40 dB. Regions of measurement of the Prony parameters wherein the partial amplitude-to-noise ratio is lower than 14 dB must be excluded from consideration in this case.

ANALYSIS OF THE EXPERIMENTAL RESULTS

The proposed processing technique has been tested in a study of the temperature dependence of the photoconductivity relaxation curves of high-resistance silicon obtained by gold doping.

Each relaxation curve has been processed by the Prony method, which discloses the presence of two damped exponential functions in each curve [Eq. (3) with $N=2$]. Figure 4 shows the temperature dependence of the parameters of these two exponentials: the partial relaxation times $\tau_{\text{fast}}(T)$ and $\tau_{\text{slow}}(T)$ along with the corresponding amplitudes $A_{\text{fast}}(T)$ and $A_{\text{slow}}(T)$.

To determine the errors of the Prony parameters, we have taken into account experimental errors together with processing errors. We use the simulation results (Figs. 2 and 3) to estimate the latter. Some of the experimental points on the temperature curve of τ_{slow} are not shown, since the reliability of these points is low on account of the smallness of the partial amplitude A_{slow} in the corresponding temperature range (Fig. 4, 140 K$<T<$200 K).

In the temperature range where the τ_{slow} rises abruptly [$\sim M/N_{CM}$, Eq. (2)] the amplitude $A_{\text{slow}}(T)$ decays by the same law ($\sim N_{CM}/M$). It is readily shown that this situation is possible if the α-attachment mechanism is operative. The situation is physically straightforward: The lower the temperature, the smaller is the thermal excursion from the at-

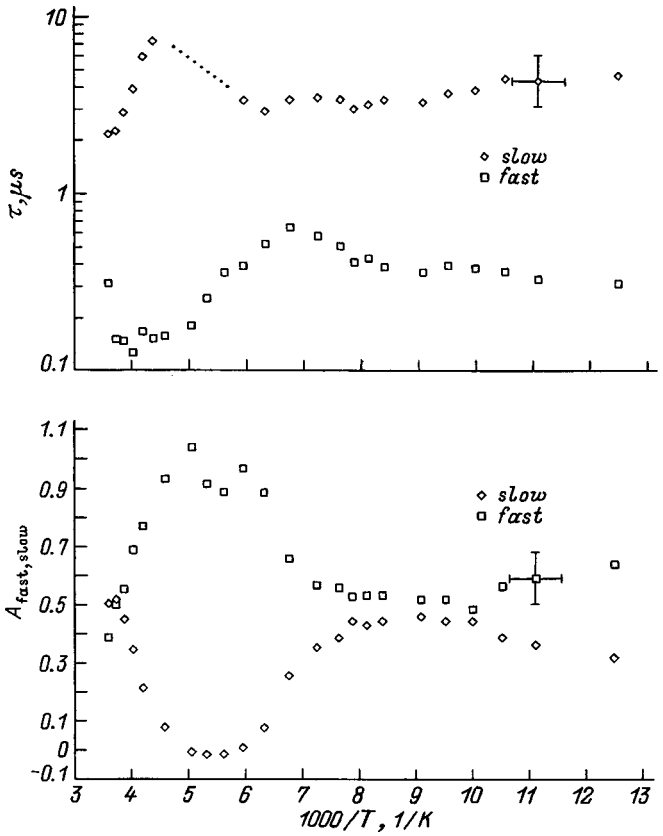

FIG. 4. Results of processing the experimental data. Results of processing a family of photoconductivity decay relaxation curves by the Prony method. A gold-doped silicon sample is investigated.

tachment level, and the smaller is the fraction of nonequilibrium electrons residing a long time at the level (τ_{slow} increases). On the other hand, as the temperature is lowered, the decrease in the thermal excursion means that a smaller fraction of electrons in the conduction band is required to maintain equilibrium with the attachment level, and A_{slow} therefore decreases as the temperature is lowered. The attachment energy level determined from the dependence $\tau_{\text{slow}}(1/T)$ or $A_{\text{slow}}(1/T)$ is equal to $[0.17\pm0.01]$ eV, which coincides with the energy of the acceptor oxygen-vacancy *A*-complex.[12] The fast component of the relaxation time in this temperature range corresponds to the initial trapping of electrons at free attachment levels $[\tau_{\text{fast}}=(\gamma_{cn}M)^{-1}]$. An order-of-magnitude estimate gives the density of these levels M greater than or equal to 10^{16} cm^{-3}. For an energy of 0.17 eV and a density of states for silicon $N_c\approx 2.9\times10^{19}$ cm^{-3} (at 300 K) the recombination coefficient at the attachment level γ_{cn} is at most equal to 5×10^{-8} cm$^3\cdot$s, which is reasonable for the trapping of an electron in a neutral acceptor state of the attachment level.

In analyzing these data, it is important to recall that the exciting irradiation time is shorter than τ_{fast} over the entire temperature range, so that the filling of the attachment levels does not reach saturation during irradiation; once the external excitation stops, the levels only begin to fill, trapping nonequilibrium carriers from the band. In this way our assumed conditions differ from those usually discussed,[10] so that the relations between A_{fast} and A_{slow} for α-attachment can vary with the temperature as shown in Fig. 4, in which case

$$A_{\text{fast}}\sim\Delta n(0)\frac{M}{N_{CM}+M},\quad A_{\text{slow}}\sim\Delta n(0)\frac{N_{CM}}{N_{CM}+M},$$

where $\Delta n(0)$ is the density of nonequilibrium electrons at the initial time; since $A_{\text{fast}}>A_{\text{slow}}$ in the investigated temperature range (Fig. 4), we have $M\geqslant N_{CM}$.

If we stay within the confines of the given model, the emergence of A_{slow} (and, accordingly, τ_{slow}) at lower temperatures can be attributed to narrowing of the energy gap between the acceptor complex and the conduction band [see Eqs. (2); the condition $(M/N_{CM})\gg1$ fails as N_{CM} increases], and the temperature dependence of τ_{slow} in the temperature range $T<180$ K is the temperature dependence of the recombination coefficient γ_n at the level governing the recombination time in the α-attachment model.[10] In this same temperature interval the temperature dependence of τ_{fast} is determined by the temperature dependence of the recombination coefficient of the attachment level.

Another explanation for this phenomenon could be the activation of a restrictive mechanism of recombination on the surface[9] with a characteristic time τ_s for large τ_{slow}. This time, in turn, is either approximately equal to $\tau_s=b^2/(16D)$ when recombination is restricted by the influx of nonequilibrium carriers to the surface, as can happen when the surface recombination rate s exceeds the threshold value $s_{\text{thr}}=D\pi/b$ (in our case $s_{\text{thr}}\geqslant300$ cm/s for $b=0.1$ cm and $D\geqslant10$ cm^2/s), or it is approximately equal to $\tau_s=b/(2s)$ in the opposite case $s\leqslant s_{\text{thr}}$. For the first case, accordingly, we have $\tau_s\leqslant10$ μs, which is close to the experimental time limit (Fig. 4), and in the second case for $\tau_s=3\mu$s (Fig. 4) we obtain $s\approx3\times10^4$ cm/s, which contradicts the condition $s\leqslant s_{\text{thr}}\approx300$ cm/s. We can therefore conclude that the first alternative holds in our case, and the time limit is specified as $\tau_s=b^2(16D)$. The temperature dependence of $\tau_{\text{slow}}\approx\tau_s$ is then determined by the temperature dependence of the diffusion coefficient $D=kT/e\cdot\mu(T)$, which at the point of maximum mobility $\mu(T)$ (≈100 K for silicon) has the form T^{ν}, $0<\nu<1$, so that $\tau_s\sim1/T^{\nu+1}$ decreases as the temperature increases, as is indeed observed experimentally for τ_{slow} (Fig. 4). The amplitude ratio in this situation describes the ratio of the fractions of the nonequilibrium density of nonequilibrium carriers that vanish as a result of surface recombination (slow component) and at bulk recombination levels (fast component).

Another important consideration is that only the electronic component of the signal is addressed here, i.e., the photoconductivity is monopolar (it must be assumed that $\tau_n\gg\tau_p$). If this were not the case, the amplitude ratio would more aptly reflect the ratio of the electron (μ_n) and hole (μ_p) mobilities, because the reflected power signal is proportional to the conductivity. The satisfaction of the condition $\tau_n\gg\tau_p$ is confirmed by the large asymmetry with which the levels are filled and by the large asymmetry of the capture cross sections of the acceptor state for gold in *n*-type silicon.[11] We must naturally conclude, therefore, that the hole lifetime τ_p is very short, not exceeding the excitation irradiation time.

To determine the errors of the Prony parameters, we have taken into account the experimental errors together with the errors of the processing method. We have estimated the latter on the basis of the simulation results (Figs. 2 and 3).

CONCLUSION

We have demonstrated the effectiveness of a generalized Prony method for analyzing the photoconductivity decay relaxation curves in the presence of noise for a bounded bit width. Our simulation has shown that the Prony method yields satisfactory results when the partial amplitude-to-noise ratio is at least 14 dB.

The method has been tested in an analysis of the photoconductivity decay curves in gold-doped silicon irradiated with laser pulses in a microwave field. An analysis of the experimental data for the investigated sample has enabled us to disclose the onset of an entrained α-attachment level situated at a depth of 0.17 ± 0.01 eV; this energy is close to the activation energy of the *A*-center oxygen-vacancy complex (0.16 eV).

The method can be used to systematically trace the temperature variations of both the relaxation times and the corresponding partial amplitudes, thereby opening additional information channels for the acquisition of data on the evolution of recombination processes.

[1] Yu. G. Amirov, A. M. Danilevskiĭ, and V. E. Chelnokov, Fiz. Tekh. Poluprovodn. **10**, 1986 (1976) [Sov. Phys. Semicond. **10**, 1186 (1976)].
[2] V. V. Grigor'ev, V. V. Zuev, M. M. Mekhtiev *et al.*, Fiz. Tekh. Poluprovodn. **24**, 2031 (1990) [Sov. Phys. Semicond. **24**, 1262 (1990)].
[3] É. I. Zavaritskaya and V. V. Kudinov, Fiz. Tekh. Poluprovodn. **18**, 2160 (1984) [Sov. Phys. Semicond. **18**, 1347 (1984)].
[4] Yu. A. Anoshin, V. M. Bazin, and A. S. Darevskiĭ, Zh. Tekh. Fiz. **59**(6), 165 (1989) [Sov. Phys. Tech. Phys. **34**, 688 (1989)].
[5] V. V. Zuev, S. I. Rasmagin, G. M. Voronkova, and Yu. A. Voronov, in *Conference on the Reliability of Quality Control of Electronics Products* [in Russian], Sevastopol, 1990.
[6] S. L. Marple Jr., *Digital Spectral Analysis with Applications* (Prentice-Hall, Englewood Cliffs, N.J., 1987; Sevastopol, 1990).
[7] I. S. Berezin and N. P. Zhidkov, *Computing Methods*, Vols. 1 and 2 (Pergamon Press, New York-Oxford, 1965; Nauka, Moscow, 1959).
[8] A. E. Mudrov, *Numerical Methods for the Personal Computer* [in Russian], Tomsk, 1992.
[9] V. L. Bonch-Bruevich and R. G. Kalashnikov, *Physics of Semiconductors* [in Russian], Moscow, 1977.
[10] S. M. Ryvkin, *Photoelectric Effects in Semiconductors* (Consultants Bureau, New York, 1964; Fizmatgiz, Leningrad, 1963).
[11] B. R. Levin, *Theoretical Foundations of Radiation Physics*, Vol. 2 [in Russian], Moscow, 1968.

Translated by James S. Wood

Energy barriers at the interfaces in the MIS system Me–Yb_2O_3–Si

V. A. Rozhkov and A. Yu. Trusova

Samara State University, 443011 Samara, Russia
(Submitted December 3, 1997)
Zh. Tekh. Fiz. **69**, 60–64 (April 1999)

The electrophysical properties of silicon MIS structures with ytterbium oxide as the insulator are investigated. It is established that the electrical conductivity of the MIS structures are described by a Poole–Frenkel mechanism. The capacitance–voltage curves are used to measure the trapped charge in the insulator and the density of surface states. The energy barriers for electrons at the interfaces are determined by the method of internal photoemission of charge carriers into the insulator. The parameters of deep electron traps in ytterbium oxide are investigated. © *1999 American Institute of Physics.* [S1063-7842(99)01104-6]

INTRODUCTION

The extensive application of metal–insulator–semiconductor (MIS) systems in semiconductor discrete devices and integrated circuits has created a need to search for promising new dielectric materials and to investigate their properties. Of considerable scientific and practical interest are rare-earth oxides, whose films have high dielectric constants ($\varepsilon = 8-20$) and high resistivities ($\rho = 10^{12}-10^{15}\,\Omega \cdot$cm) and are heat-resistant and chemically stable.[1,2] Investigations have shown that these materials can be used to fabricate silicon MIS varicaps with a high capacitance overlap factor,[3] memory electric switches,[4] and effective brightening and passivating coatings for silicon photoelectric devices.[5] However, the energy barriers at the phase interfaces and the energy band diagrams of silicon MIS systems utilizing insulator films made from oxides of RE elements have not been determined to date. We have therefore investigated the photoelectric properties of silicon MIS structures using ytterbium oxide as the insulator, and we have determined the energy barriers at the interfaces and the parameters of electron trapping centers in the insulator.

SAMPLE PREPARATION AND MEASUREMENT PROCEDURE

MIS structures of the type Me–Yb_2O_3–Si (Me stands for Al or Ni) were prepared on KÉF-5 (111)-oriented and KDB-4.5 (100)-oriented single-crystal silicon substrates. Prior to deposition of the insulator film the polished silicon wafers were degreased by boiling in toluene, followed by ultrasonic washing in acetone. The native oxide layer was removed by processing in a dilute aqueous solution of hydrofluoric acid. An ytterbium oxide film of thickness $d = 0.25-0.28\,\mu$m was prepared by the thermal spraying of metallic ytterbium from a molybdenum boat in a 10^{-5} Torr vacuum on a VUP-5 apparatus with subsequent thermal oxidation of the resulting Re layer at a temperature of 500–550 °C for 30 min. Semitransparent electrodes of area 0.4 mm^2 deposited on the Yb_2O_3 film by the thermal spraying of aluminum or nickel through a stencil in a $(1-2)\times 10^{-5}$ torr vacuum. A solid aluminum contact was deposited on the opposite polished surface of the silicon substrate.

The current–voltage I–V curves and current–temperature curves were measured at a constant voltage by means of an É D-05M electrometer. The density of surface states N_{SS} at the Yb_2O_3–Si interface and the fixed charge in the insulator Q_F were calculated by the method of high-frequency (1 MHz) capacitance–voltage (C–V) curves. The energy barriers at the metal–Yb_2O_3 and Si–Yb_2O_3 interfaces were determined by the method of internal photoemission of electrons into the insulator[6,7] from the spectral and voltaic dependences of the photocurrent. The photoemission currents were measured by means of an electrometer with a sensitivity of $\sim 10^{-15}$ A during illumination of the samples by monochromatic light obtained from a DKsÉL 1000-5 1-kW xenon lamp and an MDR-2 high-transmission monochromator.

RESULTS AND DISCUSSION

The I–V curves of the Al–Yb_2O_3–Si structures (Fig. 1) for positive (+ on Me) and negative voltage polarities are linear in the coordinates $\{\ln(I\cdot d/V), V^{1/2}\}$, where I is the current and V is the applied voltage. The rectification ratio for the various samples was in the interval 1.2–2.0. The resistivities of the ytterbium oxide films, determined from the I–V curves, were equal to $10^{14}-10^{15}\,\Omega\cdot$cm. The current–temperature curves, measured in the temperature range $T = 290-400$ K for various values of the dc voltage on the structure, become linear in the coordinates $\{\ln I, 10^3/T\}$ (Fig. 2). The current activation energy in temperature units, calculated from these curves, decreases monotonically from 1.1–1.3 to 0.86–0.9 eV as the voltage is increased from 2 V to 5 V. These results show that the electrical conductivity of the MIS structures is limited by the volume of the insulator and is satisfactorily described by the Poole–Frenkel mechanism.

The C–V curves of the Al–Yb_2O_3–Si structures (Fig. 3) have a typical high-frequency profile with saturation of the capacitance upon accumulation and inversion at the surface of the semiconductor. The capacitance of the samples decreases approximately from 160 pF to 40 pF as the dc voltage is varied from accumulation values to values corresponding to inversion at the surface of the semiconductor. The control

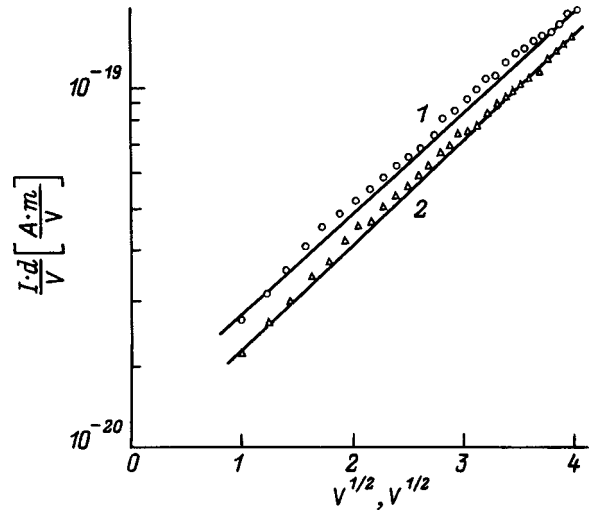

FIG. 1. Current–voltage curves of Al–Yb_2O_3–Si structures for positive (*1*) and negative (*2*) voltages.

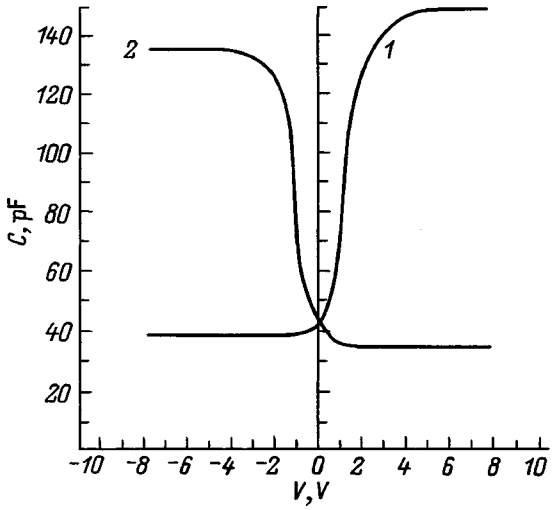

FIG. 3. Capacitance–voltage curves of Al–Yb_2O_3–*n*-Si (*1*) and Al–Yb_2O_3–*p*-Si (*2*).

voltages lie between -2 V and 2 V. The flat-band potential and the capacitance overlap factor for various samples are equal to 0.5–0.9 V and 3–4, respectively. The hysteresis of the C–V curves along the voltage axis falls within the interval 0.2–0.5 V. The density of surface states and the trapped charge in the insulator are equal to $(2.5-4)\times 10^{11}\,\mathrm{cm}^{-2}\cdot\mathrm{eV}^{-1}$ and $(4-7)\times 10^{-8}\,\mathrm{C/cm}^2$, respectively.

Figure 4 shows typical plots of the photocurrent per photon $I_f^{1/n}$ as a function of the photon energy $h\nu$ of an Al–Yb_2O_3–Si structure in the coordinates $\{I_f^{1/n}, h\nu\}$ for various positive and negative voltages; $n=2$ and $n=3$ correspond to the application of negative and positive voltages, respectively. It is evident that $I_f^{1/n}$ increases linearly as the photon energy increases in an interval $\sim 10\,kT$. The curves for a positive voltage on the electrode attached to the oxide

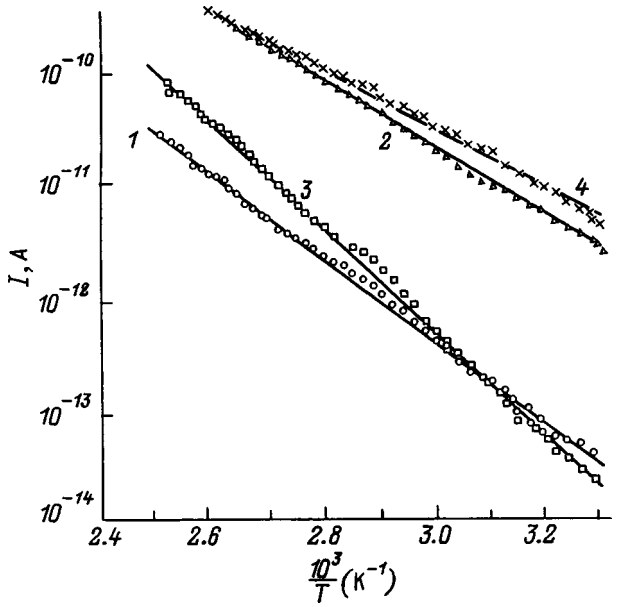

FIG. 2. Temperature dependence of the current in an Al–Yb_2O_3–Si structure for various applied voltages: (*1*) $V=-2$ V; (*2*) -5 V; (*3*) $+2$ V; (*4*) $+5$ V.

(Fig. 4a) corresponds to the photoemission of electrons from silicon. The emission of holes from the metal is negligible in this case, because the threshold photon energy for this polarity does not depend on the material of the field electrode. Moreover, the C–V curves show that the capture of negative charge by deep traps in Yb_2O_3 is observed as the photocurrent passes through the insulator in this case. The threshold photon energy corresponds to the energy E_b of the potential barrier at the Si–Yb_2O_3 interface for the given applied voltage.

The curves for negative biases on the field electrode (Fig. 4b) correspond to electron emission from the metal. The threshold photon energy for this polarity depends on the material of the field electrode, and a negative charge builds up in the insulator layer. The threshold photon energy for these curves is equal to the energy of the potential barrier at the metal–ytterbium-oxide interface for the given voltage.

Investigations have shown that the dependence of the potential barrier E_b on the square root of the applied voltage is linear, and the photocurrent for electric fields in the insulator greater than 10^5 V/cm is described by the expression

$$I_t = A(h\nu - E_{b0} + kV_d^{1/2})^n, \tag{1}$$

where E_{b0} is the height of the potential barrier at zero electric field in the insulator, k is the Schottky coefficient, V_d is the voltage drop across the insulator, and A is a coefficient that depends on $h\nu$.

The zero-voltage energy barriers at the boundary of the ytterbium oxide boundary can be determined by extrapolating the linear graphs of E_b versus $V^{1/2}$ to their intercepts on the E_b axis.

Figure 5 shows typical photovoltaic curves representing the voltage dependence of the photoinjection current from aluminum into the ytterbium oxide film in the coordinates $\{I_f^{1/2}, V^{1/2}\}$ for two different photon energies. The curves acquire good linearity in these coordinates, corroborating the analytical relation given for the photoemission current. The heights of the energy barriers at the interfacial boundaries of the Yb_2O_3 film, as determined from these curves by the tech-

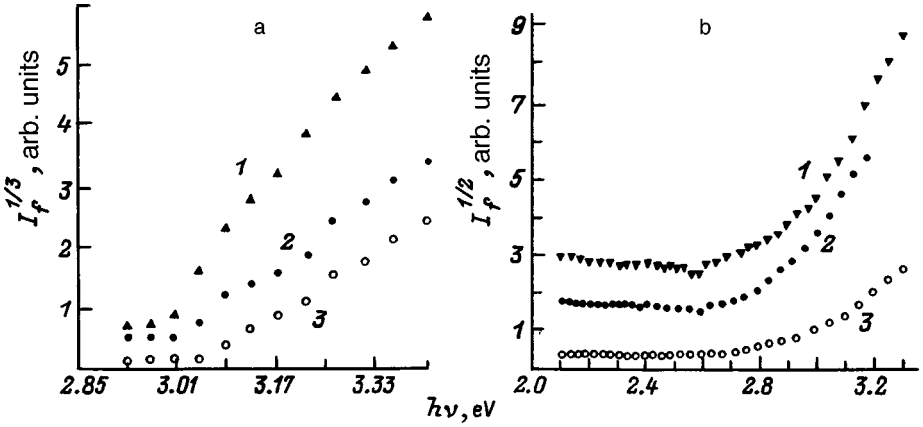

FIG. 4. Graphs of the normalized photocurrent of an Al–Yb_2O_3–Si structure as a function of the incident photon energy for (a) positive voltages: (*1*) $V=3$ V; (*2*) 4 V; (*3*) 5 V; (b) negative voltages: (*1*) $V=-2$ V; (*2*) -4 V; (*3*) -5 V.

nique described in Ref. 8, were practically equal to the values obtained from analysis of the spectral curves of the photocurrent. The heights of the energy barriers at the phase interfaces in ytterbium oxide MIS structures are shown in Table I.

The heights of the energy barriers at the interface with silicon are identical within the experimental error limits for MIS systems containing *n*-type and *p*-type silicon. The energy barrier for a Ni–Yb_2O_3 interface is 0.4 eV higher than for an Al–Yb_2O_3 interface. The resulting difference between the barrier energies is equal to the absolute difference in the work functions for the emission of electrons from the metal.

When Al–Yb_2O_3–Si structures are irradiated in the ultraviolet range with a photon energy ($h\nu=5.3$ eV) greater than the width of the Yb_2O_3 band gap ($E_g=5.2$ eV), the *C–V* curves are observed to shift along the voltage axis toward positive values for both polarities of the applied voltage, indicating the capture of electrons by deep traps in the Yb_2O_3. The graphs of the shift ΔV_{FB} of the flat-band potential as a function of the irradiation time have a sublinear profile (Fig. 6). In this case the rate of change of ΔV_{FB} increases as the voltage on the structure is increased.

After an Al–Yb_2O_3–Si structure, under the influence of an applied dc voltage, is irradiated with monochromatic light having a photon energy higher than the potential barrier but smaller than the width of the band gap of ytterbium oxide, a similar shift of the *C–V* curves along the voltage axis toward positive values is observed. This result attests to the capture of negative charge by deep traps in the insulator. In this case the dependence of the shift of the flat-band potential of the *C–V* curves on the irradiation time is sublinear, tending to saturation after long (approximately one hour) irradiation times. The trapped charge and the flat-band potential shift increase as the voltage applied to the structure is increased. The trapped charge is preserved for a long time at room temperature after the irradiation and dc voltage are discontinued. The effective charge trapped in the insulator, calculated from the shift of the *C–V* curve, is an order of magnitude smaller than the charge transferred through the structure; the latter charge is determined by integrating the current over time.

When an Al–Yb_2O_3–Si structure is irradiated with light, the photocurrent decreases with time (Fig. 7), attaining a steady-state value. The observed photocurrent decay kinetics is attributable to the photoinjection and trapping of electrons in the insulator. Charge carriers trapped in deep states in the insulator diminish the electric field in it, causing the photocurrent through the structure to decrease.

The photostimulated depolarization method has been used to determine the energy depth of the electron traps in Yb_2O_3. After photopolarization, the MIS structure was short-circuited to an electrometer without any external voltage, and the spectral curves of the depolarization photocurrent at photon energies lower than the energy barriers at the boundary of the Yb_2O_3 film were measured. The spectral curves of the depolarization photocurrent, measured immediately after photopolarization at incident photon energies of 5.2 eV for 1 h and in repeated measurements are shown in Fig. 8. It is evident that after photopolarization the indicated curves ac-

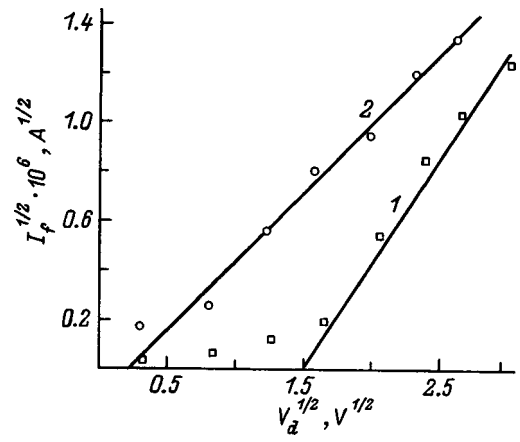

FIG. 5. Photovoltaic curves (photocurrent versus voltage) for different photon energies: (*1*) $h\nu=2.89$ eV; (*2*) 2.75 eV.

TABLE I. Interface barrier energies in MIS structures.

Interface with Yb_2O_3 phase	Height of the energy barrier, eV
Al	2.9–2.92
Ni	3.3–3.32
n-Si	3.19–3.21
p-Si	3.18–3.21

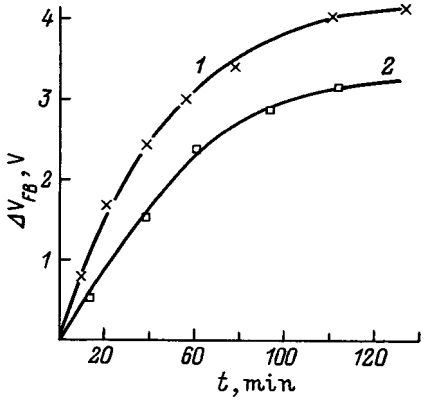

FIG. 6. Dependence of the shift of the flat-band potential on the irradiation time for different voltages: (*1*) $V=-7$ V; (*2*) 6 V.

quire a photocurrent peak that had not been observed before the insulator was charged and whose amplitude decreases in repeated measurements. This photocurrent is produced by the photoexcitation of trapped electrons from traps into the conduction band of the insulator and their drift under the influence of the built-in electric field. In repeated measurements of these curves the degree of filling of the electron traps decreases, and the amplitude of the photocurrent peak is smaller. After complete photodepletion of the electron traps the peak vanishes, and the photocurrent (curve *4* in Fig. 8) and the C–V curves recover their original (before photopolarization) values and positions. This result confirms that the capture of electrons by deep traps and photostimulated depletion of the latter take place when the structure is illuminated under certain conditions. The experiment described here is highly reproducible. The results in Fig. 8 can be used to estimate the energy depth of the electron traps in Yb_2O_3. It is evident that the electron trapping centers are situated almost in the middle of the Yb_2O_3 band gap at a distance of 2.4–2.7 eV from the bottom of the conduction band. The trapped charge densities in the insulator, calculated from the kinetic discharge photocurrent kinetic curves, lie in the interval $(1.8–1.9)\times10^{-8}$ C/cm^2.

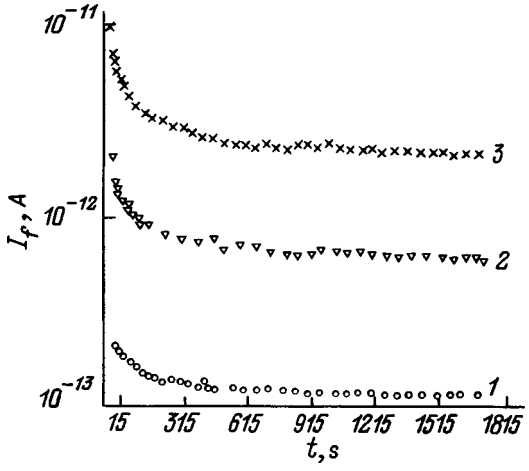

FIG. 7. Kinetic curves of the photocurrent for various voltages applied to the structure: (*1*) $V=-3$ V; (*2*) -5 V; (*3*) -6 V.

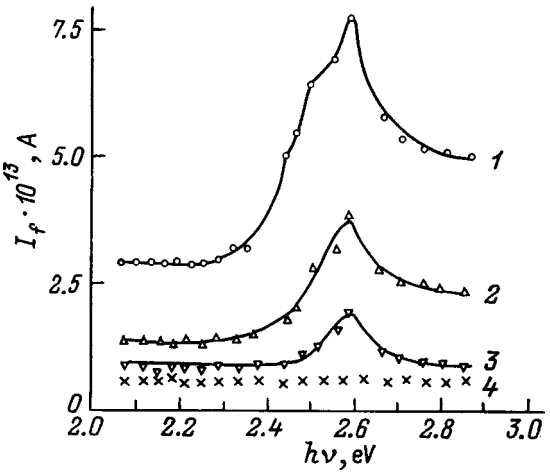

FIG. 8. Spectral curves of the photodepolarization current: (*1*) after photopolarization; (*2*) repeated measurement; (*3*) third measurement; (*4*) fourth measurement.

A simultaneous analysis of the photovoltaic and capacitance-voltage curves before and after irradiation of the structures with light leads to the assumption that the charge trapped during irradiation is located in the interior of the insulator. The magnitude and centroid of the trapped charge in the insulator have been determined from the shift of the photovoltaic curve after irradiation.[9]

Figure 9 shows the photovoltaic curves of the structure before and after irradiation with 3.1-eV photons at a dc voltage of -6 V for 40 min. After irradiation the photocurrent through the structure is observed to decrease as a result of trapped charge. The magnitude Q and centroid X_C of the trapped charge can be calculated from the relations

$$Q=\frac{\varepsilon\varepsilon_0}{d}(|\Delta V|^+ + |\Delta V|^-), \tag{2}$$

$$X_C = d\left(1+\frac{|\Delta V|^+}{|\Delta V|^-}\right)^{-1}, \tag{3}$$

where ε_0 and ε are the dielectric permittivities of vacuum and the insulator, respectively, and ΔV^+ and ΔV^- are the

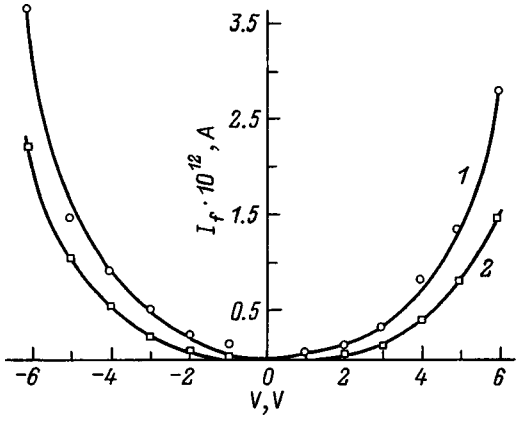

FIG. 9. Photovoltaic curves for an Al–Yb_2O_3–*n* Si structure before (*1*) and after (*2*) irradiation.

TABLE II. Parameters of electron trapping centers in an Al–Yb_2O_3–Si structure.

Structure	X_C, μm	Q_t, C/cm^2	Q_t/Q_g	Irradiation regime (injection of electrons)
Al–Yb_2O_3–*p*-Si	0.13	3.6×10^{-10}	0.16	from Al
	0.13	6×10^{-10}	0.31	from Si
Al–Yb_2O_3–*n*-Si	0.13	4.5×10^{-10}	0.21	from Al
	0.16	7.35×10^{-10}	0.12	from Si

shift of the I_F-V curves along the voltage axis due to irradiation for positive and negative polarities of the applied voltage, respectively.

The calculated values of the centroid, determined from the Si surface, of the charge Q_t trapped in the insulator and the ratio of the trapped charge to the transmitted charge Q_g for various irradiation regimes are given in Table II.

It is evident that the centroid of the trapped charge is situated almost in the center of the insulator layer. This result shows that the Yb_2O_3–Si interface is a sharp boundary in the investigated structures and does not contain a disturbed transition layer of finite thickness. This condition is probably associated with the comparatively low temperatures at which the insulator films are prepared, so that large stresses are not created in them.

[1] O. S. Vdovin, Z. I. Kir'yashkina, V. N. Kotelkov *et al.*, *Rare-Earth Oxide Films in MIM and MIS Structures* [in Russian], Universitet, Saratov, 1983.
[2] V. A. Rozhkov, O. S. Vdovin, V. N. Kotelkov, and A. M. Sverdlova, Mikroelektronika **5**, 28 (1976).
[3] V. A. Rozhkov, O. S. Vdovin, V. N. Kotelkov, and A. M. Sverdlova, Élektron. Promst., No. 1, 22 (1973).
[4] V. A. Rozhkov and A. I. Petrov, Pis'ma Zh. Tekh. Fiz. **11**, 49 (1985) [Sov. Tech. Phys. Lett. **11**, 21 (1985)].
[5] Yu. A. Anoshin, A. I. Petrov, V. A. Rozhkov *et al.*, Zh. Tekh. Fiz. **64**(10), 118 (1994) [Tech. Phys. **39**, 1039 (1994)].
[6] B. E. Deal, E. H. Snow, and C. A. Mead, J. Phys. Chem. Solids **27**, 1873 (1966).
[7] R. J. Powell, J. Appl. Phys. **40**, 1093 (1969).
[8] R. J. Powell, J. Appl. Phys. **41**, 2424 (1969).
[9] D. J. Di Maria, J. Appl. Phys. **47**, 4073 (1976).

Translated by James S. Wood

Efficiency of stimulation of gaseous chemical reactions by a diffracted light beam whose energy is absorbed by vibrational transitions

O. A. Gunaze and V. A. Trofimov

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted November 25, 1996; resubmitted February 2, 1998)
Zh. Tekh. Fiz. **69**, 65–73 (April 1999)

Laser-induced reversible chemical reaction kinetics in a gas mixture with the absorption of light energy by molecular vibrational transitions is investigated by mathematical modeling. The influence of beam diffraction, the relaxation time of vibrational energy into heat, and the order of vibrational transition on the structure of the region of high absorption and on the propagation of switching waves is analyzed. The discussion applies to pulses having a duration much longer than the time of energy transfer between vibrational levels. © *1999 American Institute of Physics.* [S1063-7842(99)01204-0]

INTRODUCTION

Laser-induced processes in chemically active gaseous media are major objects of investigation by virtue of their many practical applications. However, the control of chemical reaction in a gas is itself a typical example of problems that can be solved in a vigorously growing science: laser dynamics. The implementation of various nonlinear optical self-organization processes, including optical bistability, in the system comprising the light beam and chemically active gas mixture is attracting the attention of many researchers.[1–11] In papers devoted to this topic, as a rule, the diffraction of optical radiation has either been ignored altogether or is taken into account in the aberration-free approximation, which is valid only in the initial stage of interaction of optical radiation with a chemically active gas mixture, i.e., while nonlinear absorption can still prevent its profile from being transformed into a structure with two maxima. In the meantime, when the input radius of the beam is decreased, the influence of its diffraction can fundamentally alter the space–time evolution of the region of high concentration of the reaction product.[8]

It must also be borne in mind that diffraction effects occur even when the initial diffraction length of the beam is several times the length of the cell containing the gas mixture, because the concentration of the reaction product increases rapidly during the interaction of laser radiation with a chemically active gas mixture under certain conditions (optical bistability conditions) in the regions of highest intensity in the medium. As a result, owing to the dependence of the absorption on the concentration of the given species of gas, light energy is ''sapped'' in this region of the beam. A consequence of this phenomenon is a partitioning of the beam into subbeams, for each of which the diffraction length decreases and, hence, the same happens for the whole beam.

Another meagerly studied problem is the stimulation of chemical reactions when light energy is injected through molecular vibrational levels and is subsequently converted into thermal energy. This technique for the control of reactions is conducive to the more efficient utilization of laser energy:[9] A high concentration of reaction product can be attained at a lower temperature of the medium and at a higher beam output power. We also emphasize that the interaction of optical radiation with molecules with the absorption of its energy by a vibrational transition, in turn, exhibits the bistability property under certain conditions.[3] The coexistence of bistable dependences in several variables can also qualitatively change the kinetics of a laser-induced gas reaction.

In this paper we analyze the interaction of a diffracted laser beam with a two-component gas mixture when light energy is absorbed by a molecular vibrational transition.

STATEMENT OF THE PROBLEM

We wish to investigate the propagation of optical radiation in a cell containing a gas mixture; the cell is closed to mass transfer. Let heat transfer with the environment take place through the side wall of the cell, and let the end walls be thermally insulated. We also assume that energy is transferred between molecular vibrational levels in a time much shorter than the pulse duration. As a result of this assumption, the populations of the vibrational levels fit a Boltzmann distribution.[3] However, the transfer time exceeds the relaxation time of vibrational energy into translational energy. The variation of the temperature of the medium must therefore be taken into account. We also assume that the molecular vibrational levels are not equidistant and that the absorption of light energy is a resonance process in the $n \rightarrow n+1$ transition.

We consider the influence of a light beam having a radius such that the diffraction length is commensurate with the length of the cell containing the gas mixture. This condition is indicative of the need to take optical diffraction into account. We also assume that the length of the cell is comparable with the path of nonlinear refraction under the influence of the thermal lens due to heating of the medium and also the influence of the thermal lens due to the variation in the composition of the medium as a result of chemical reaction. We assume in this regard that the concentrations of the reaction product and the starting mixture are locally related.

This approximation is valid, for example, in a reaction $A+B\leftrightarrow C$ (where A and B are the starting substances, and C is the reaction product) if the concentration of one of the starting substances (say B) is highly predominant over the other substances and if the diffraction coefficients of the reaction product and the initial gas differ only slightly. The chemical reaction can then be characterized by a single function: the concentration of its product. In this case the interaction of the diffracted beam with the medium can be described by the system of dimensionless equations

$$\frac{\partial A}{\partial z}+iD\Delta_\perp A+i\varepsilon_{nl}A+\delta_0\delta(N,E)A=0,$$

$$0<z\leqslant L_z,\quad 0<x<L_x,$$

$$\varepsilon_{nl}=\alpha(T-T_0)+\beta(N-N_0),\quad \delta(N,E)=\sigma N\frac{E^n}{(1+E)^{n+2}},$$

$$\varepsilon\frac{\partial T}{\partial t}=\chi\Delta_\perp T+\gamma f(N,T)+T_E\frac{E-E(T)}{\tau_{VT}},\quad t>0,$$

$$\frac{\partial E}{\partial t}=qI\delta(N,E)-\frac{E-E(T)}{\tau_{VT}},$$

$$\frac{\partial N}{\partial t}=f(N,T)+D_N\Delta_\perp N,\quad \Delta_\perp=\frac{\partial^2}{\partial x^2},$$

$$f=(1-N)\exp\left(-\frac{1}{T}\right)-kN\exp\left(-\frac{T_c}{T}\right),\quad I=|A|^2,\qquad (1)$$

subject to the initial and boundary conditions

$$A|_{z=0}=A_0(x)\sqrt{I_0(t)},$$

$$I_0(t)=(1-\exp(-t/\tau))^2,\quad \tau=0.1,$$

$$A_0(x)=\exp\left(-\frac{(x-L_x/2)^2}{a^2}\right),\quad A|_{x=0,L_x}=0,$$

$$\left.\frac{\partial N}{\partial x}\right|_{x=0,L_x}=0,\quad \chi\left.\frac{\partial T}{\partial x}\right|_{x=0,L_x}=\pm\eta(T-T_0),$$

$$A|_{t=0}=0,\quad T|_{t=0}=T_0,\quad f|_{t=0}=0,\quad E|_{t=0}=E(T_0).\qquad (2)$$

In Eqs. (1) and (2) the complex amplitude $A(z,x,t)$ is normalized to its maximum input value, $N(z,x,t)$ is the dimensionless concentration of the reaction products, $T(z,x,t)$ is the temperature of the medium, normalized to the activation energy of the direct reaction in temperature units, $E(z,x,t)$ is the normalized vibrational energy per unit volume of the gas in the given mode, and $E(T)$ is the equilibrium value of this energy at the specified temperature T.

Under the above-stated assumptions the equilibrium vibrational energy is given by the expression

$$E(T)=\frac{1}{\exp(T_E/T)-1}.$$

In the calculations we use an expression for $E(T)$ prevalent in the literature:[3] $E(T)=\exp\{-(T_E/T)\}$. Moreover, z is measured in units of the diffraction length $l_d=2ka_0^2$, k is the wave number, a_0 is the initial beam radius, the transverse coordinate x is measured in units of a_0, and D is a coefficient, which is equal to unity under the given normalization, but is retained in Eqs. (1) to aid in analyzing the influence of beam diffraction on the spatial distributions of the temperature, concentration, and intensity of the light beam.

The coefficient δ_0 characterizes the amplitude absorption of light by all molecules associated with the reaction product. The time t is measured in units of the characteristic relaxation time. The parameter ε characterizes the rate of change of the temperature relative to the rate of the change of the composition of the gas mixture. The coefficients χ and D_N are equal to the ratios of the characteristic relaxation time to the characteristic heat-conduction and diffusion times, respectively. The parameter q is the ratio of the initial beam power to the power spent in changing the composition of the mixture from the original gas to the reaction product. The coefficient γ, which is proportional to the term (T_c-1), characterizes the reaction energy, T_c is the ratio of the activation energies of the reverse and forward reactions, k is the ratio of the pre-exponential factors of the reverse to the forward reaction, τ_{VT} is the normalized time constant for the relaxation of vibrational to translational energy, σ is the total absorption cross section of the maximum number of molecules of a given species in the vibrational transition $n\rightarrow n+1$ with the difference between the populations of levels n and $n+1$ of the absorbing mode as specified by the ratio $E^n/(1+E)^{n+2}$, and the parameter T_E is equal to the ratio of the photon energy to the activation energy of the forward reaction.

Self-induced effects of the light beam are described by the e_{nl} term in Eqs. (1), in which α is the ratio of the input beam power P_0 to the thermal self-focusing power ($\alpha<0$ corresponds to self-defocusing of the beam), and β is proportional to the difference in the refractive indices of the reaction product and the starting substance. It is positive if the starting substance is less dense than the reaction product, and it is negative in the opposite situation; L_x is the boundary of the region along the x axis. Under the boundary conditions (2) the constant η dictates the rate of heat transfer with the environment, whose temperature is T_0.

In the numerical simulations a symmetric differencing scheme is used for the quasi-optical equation, approximating the differential operator with second-order accuracy in the space variables and time. Implicit differencing schemes of first-order approximation in time and second-order approximation in the space coordinates are used for the equations in N, E, and T.

RESULTS OF THE NUMERICAL SIMULATIONS

In keeping with the subject set forth in the title to the article, we fix the interaction parameters

$$D=0.01,\quad \alpha=\beta=0,\quad \gamma=0,\quad T_E=1,$$

$$D_N=\chi=10^{-5},\quad T_c=0.375,\quad k=0.5,$$

$$a=0.1,\quad \eta=2,\quad L_x=2,\quad L_z=1,\qquad (3)$$

TABLE I. Dependence of the time of onset of efficient reaction t_{exp} on the ambient temperature T_0 with the absorption of light energy by the $2\rightarrow3$ transition ($n=2$) for $q=50$, $\sigma=10$, and $\delta_0=10$.

T_0	t_{exp}
0.25	1
0.2	10
0.18	30
0.135	150
0.125	>200

for which optical bistability is achieved in the case of weakly diffracted beams ($D\rightarrow0$). The diffraction coefficient in Eqs. (3) corresponds to the effective manifestation of diffraction along the investigated propagation channels, because the degree of its influence is governed by the ratio D/a^2.

The values of the parameter $n=0, 1, 2$ correspond to the absorption of light energy between the molecular vibrational levels 0 and 1, 1 and 2, and 2 and 3, respectively.[3] Since the number of molecules existing in the unperturbed medium in the state with quantum number j corresponding to a vibrational energy E_j fits a Boltzmann distribution, as characterized by the function $E(T)$ in Eqs. (1), the temperature of the thermostat (surrounding the gas cell) must be raised to achieve efficient stimulation of the reaction for $n=2$. Consequently, the kinetic laws for a laser-induced reversible chemical reaction with the absorption of light energy by the vibrational ground level ($n=0$) or by the first excited level ($n=1$) are analyzed for $T_0=0.125$ (as in our earlier papers), whereas for $n=2$ the numerical simulations are carried out with $T_0=0.25$. The values given below for σ and q in Eqs. (1) differ for different numbers n for the same reason. However, for the analysis to be complete we investigate the dependence of the time of onset of efficient reaction on the ambient temperature for $n=2$ with light energy absorbed in the vibrational transition $2\rightarrow3$ ($n=2$). The modeling results are shown in Table I. Here t_{exp} is interpreted as the time of onset of reaction near the input cross section of the cell. It is evident from the table that t_{exp} increases rapidly as the temperature of the thermostat is lowered. At $T_0<0.13$ an efficient reaction is not observed in our investigated time interval $0\leqslant t\leqslant L_t=200$. Outgoing heat transfer prevails over the input of light energy in this case.

a) Case $n=0$

We are concerned primarily with the laws governing the formation of the region of high concentration of the reaction product when the radiated laser energy is absorbed by the ground molecular vibrational transition. In this case the absorption coefficient decreases monotonically as the vibrational energy increases, corresponding to saturation of the vibrational transition.

It is important to stress the fact that the spatial distribution of the concentration of the reaction product and its dynamics are acted upon by several different physical factors at once. For example, an increase in the vibrational energy leads to the above-indicated saturation of the absorption coefficient and, hence, a reduction in the injection of light energy into the medium. However, owing to the large relaxation of energy into heat, the same increase in energy leads to saturation of the temperature of the gas mixture, and the manifestation of this process is also twofold. On the one hand, an increase in the temperature stimulates reaction, so that the concentration of the reaction product and, hence, the absorption coefficient increase. On the other hand, the equilibrium vibrational energy increases, reducing the fraction of its conversion into translation energy and bleaching of the medium. The end result of this behavior is slowing of the chemical reaction. Consequently, the ratio of the parameters ε and τ_{VT} assumes an important role when the input of light energy takes place through molecular vibrational transitions. It essentially determines the time of departure of the vibrational energy from its equilibrium value and is a measure of the negative feedback between the absorption of light energy and variation of the concentration of the reaction product; an increase in N causes E to increase, and this, in turn, diminishes the injection of light energy into the medium.

Finally, we discuss the effects of optical diffraction. It also has a twofold outcome. On the one hand, being manifested specifically as an increase in the beam radius, diffraction shortens the region of efficient reaction along the z axis and decreases the part of the region with the maximum concentration of the reaction product. On the other hand, diffraction can induce focusing of the light beam as a result of the refraction of its rays by the induced distributed amplitude lattice of the absorption coefficient — the reflection of rays from the less dense side regions of the medium. This phenomenon further stimulates chemical reaction in the region around the axis. The time of influx of light energy due to focusing of the beam depends on the velocity of propagation of the transverse switching wave; this velocity, in turn, depends on τ_{VT}. It obviously vanishes if the switching wave escapes from the region occupied by the beam. We note that beam focusing is also induced by the transformation of its profile to a distribution with an intensity trough on the axis as a result of the efficient absorption of light energy. It also has a finite lifetime, which is dictated by diffusion processes and thermal conductivity. These effects result in complex, time-dependent spatial distributions of the light energy, the temperature of the medium, and the concentration of the reaction product.[12]

We therefore reemphasize that the stimulation of chemical reactions in gases by diffracted light as its energy is absorbed by the ground vibrational transition is influenced by many physical factors, which simultaneously stimulate (positive feedback) and limit (negative feedback) the reaction process. We therefore begin our analysis of the interaction of a laser beam with a chemically active gas mixture by examining the influence of beam diffraction, because in the absence of an unambiguous conclusion as to what role it plays, the results of the numerical simulations can be interpreted in different ways. To that end, we carry out the calculations with the coefficient $D=10^{-5}$, for which diffraction effects have scarcely any effect on our investigated paths.

As an illustration, Fig. 1 shows plots of the time dependence of the light intensity for weak beam diffraction $D=10^{-5}$ (a) and for fully developed diffraction $D=10^{-2}$ (b). A comparison of these figures shows that for the inter-

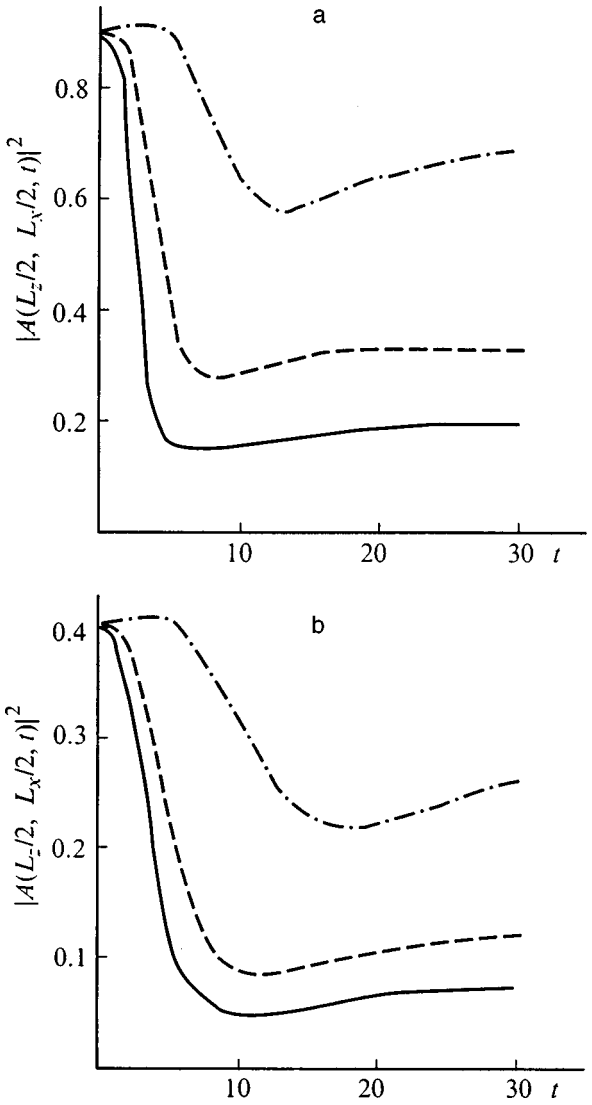

FIG. 1. Time evolution of the light intensity at the center of the cell ($z=L_z/2$, $x=L_x/2$) during the propagation of a weakly diffracted beam, $D=10^{-5}$ (a) and a strongly diffracted beam, $D=10^{-2}$ (b) in the presence of absorption of light energy by the ground vibrational transition ($n=0$) for $q=8$, $\delta_0=10$, $\sigma=1$, and various relaxation times: $\tau_{VT}=10^{-4}$ (solid curve), 1 (dashed curve), 10 (dot-and-dash curve).

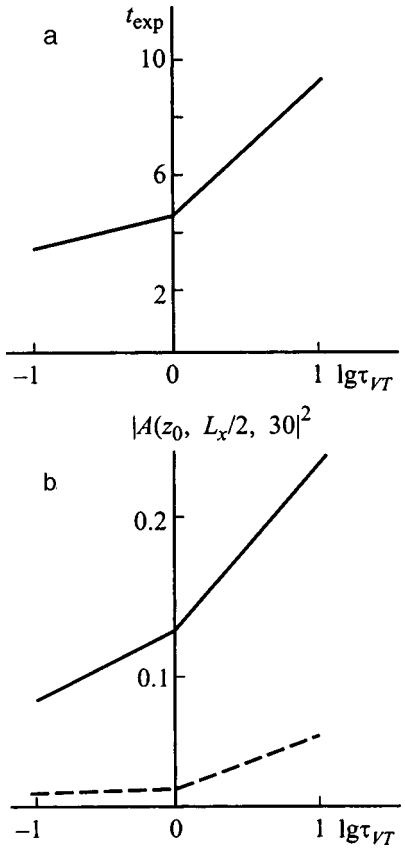

FIG. 2. (a) Dependence of the time of onset of efficient reaction on the logarithm of the relaxation time of vibrational energy at the center of the cell ($z=L_z/2$, $x=L_x/2$) in the propagation of diffracted beams ($D=10^{-2}$) for the parameters $q=8$, $\delta_0=10$, and $\sigma=1$ with the absorption of light energy by the ground vibrational transition ($n=0$); (b) intensity $I=|A(z,x,t)|^2$ on the beam axis ($x=L_x/2$) at the time $t=30$ in the center of the cell ($z=L_z/2$, solid curve) at its exit ($z=L_z$, dashed curve) as a function of the logarithm of the relaxation time (the parameters are the same as in Fig. 2a).

action parameters (3) the character of the time evolution of the intensity does not depend on the diffraction coefficient and is essentially governed by the dynamics of the absorption coefficient. However, it does have a significant influence on the attainable values of the intensity and the time at which bleaching of the medium sets in. We emphasize that, because the reaction efficiency exhibits a threshold-type dependence (due to the onset of optical bistability) on the intensity, the transition to diffracted beams is accompanied by a qualitative transformation of the region of high concentration of the reaction product, in particular, a decrease in its longitudinal length.

Several stages of interaction of the laser beam with the medium are discernible in the figure and also depend on the relaxation time of the vibrational energy: For $\tau_{VT}=10$ there is an initial interval with a slight rise in the intensity, which does not occur for $\tau_{VT}\leqslant 1$. In the initial stage of interaction the beam intensity varies only slightly, because the chemical reaction rate is insignificant.

The reaction rate then increases abruptly (explosive absorption develops as a result of the closure of positive feedback between the absorption of light energy and the increase in concentration of the reaction product). We denote this time by t_{exp}. Its dependence on the relaxation time for fully developed beam diffraction is shown in Fig. 2a.

In the next stage the absorption coefficient decreases on account of saturation of the vibrational transition, and the light intensity begins to increase gradually. A comparison of the rate of growth of the intensity for diffracted and weakly diffracted beams (Fig. 1) shows that diffraction only slightly alters the time evolution of the intensity. However, its magnitude depends to a large extent on the relaxation time and the longitudinal cross section of the cell, as illustrated in Fig. 2b. We see that the light intensity near the output cross section of the cell begins to increase for large values of τ_{VT}. This behavior is attributable to an increase in the optical thickness of the gas mixture, and bleaching of the latter requires a longer time interval for the input of light energy; this interval, in turn, depends on the relaxation time τ_{VT}. Notably, the essentially constant output intensity observed in Fig.

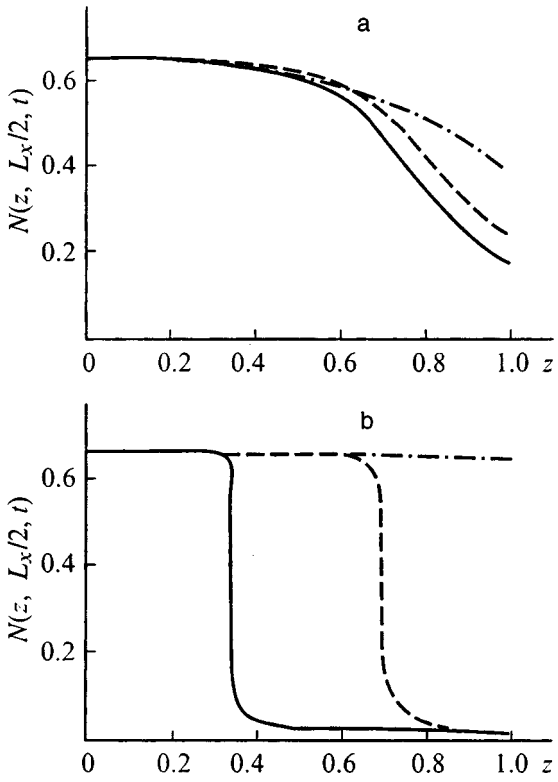

FIG. 3. Longitudinal distribution of the concentration of the reaction product on the cell axis ($x=L_x/2$) with the absorption of light energy by the ground transition $n=0$ (a) and by the first excited transition $n=1$ (b) for various relaxation times: $\tau_{VT}=0.1$ (solid curve), 1 (dashed curve), 10 (dot-and-dash curve), $D=10^{-2}$, and $\delta_0=10$ at the time $t=30$ for $q=8$ and $\sigma=1$ (a) and at the time $t=80$ for $q=50$ and $\sigma=10$ (b).

2b (dashed curve) indicates that as τ_{VT} increases, the vibrational-energy saturation of the absorption coefficient is partially offset by a longitudinal increase in the region of high concentration of the reaction product (Fig. 3a). As a result, the optical thickness of the medium remains essentially unchanged as τ_{VT} increases from 0.1 to 1.0; this fact in conjunction with the increased output intensity attests to the more efficient injection of light energy.

We now analyze the spatial distribution of the region of high concentration of the reaction product. In the case of instantaneous relaxation of the absorbed light energy into heat a very distinct longitudinal kink occurs with an essentially uniform concentration of the reaction product within the boundaries of the central region of the beam (Fig. 4a). As a consequence of high absorption the light intensity after the switching cross section is small, so that the transverse width of the region of 10% concentration of the reaction product remains constant. But if the vibrational energy relaxation dynamics becomes significant, the region of maximum concentration moves with time in both the longitudinal and the transverse direction. In this event, diffusion causes the transverse motion of the concentration wave to be superimposed with another motion induced by beam diffraction, as illustrated in Fig. 4b. This figure shows the concentration level lines, and one can clearly track the increase in the longitudinal and transverse dimensions of the region of efficient reaction in comparison with the case represented in Fig. 4a. Subsequently, at certain values of τ_{VT} the maximum concentration can also be attained at the back wall of the cell. We emphasize that the increase in the longitudinal length of the region of high concentration is accompanied by simultaneous motion of the leading edge of the region of high optical intensity along the z axis as a result of bleaching of the medium due to vibrational-energy saturation of the absorption coefficient.

A similar spatial distribution of the region of efficient reaction is encountered for weakly diffracted beams (e.g., for $D=10^{-5}$). Now, however, its transverse width decreases with increasing z, because there is no transverse flux of light energy from the beam-occupied axial zone of the cell.

We now consider the evolution of the optical intensity profile in the presence of strong diffraction ($D=0.01$). It

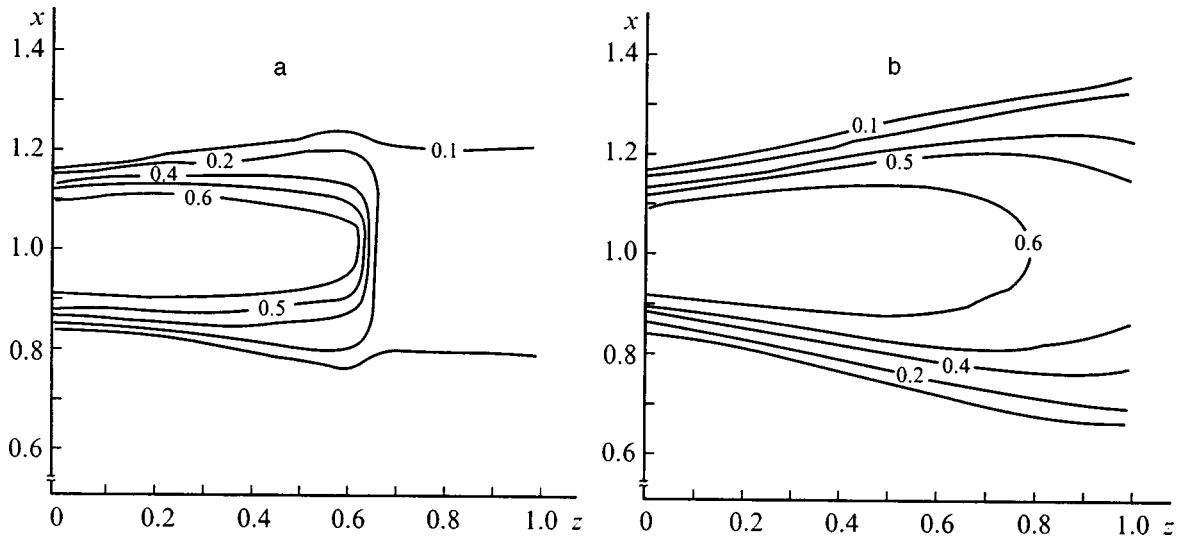

FIG. 4. Level lines of the concentration of the reaction product at the time $t=50$ for $q=8$, $\delta_0=10$, $\sigma=1$, $n=0$, $D=10^{-2}$, and relaxation times $\tau_{VT}=10^{-4}$ (a) and 10 (b).

should be noted that for all three designated relaxation times $\tau_{VT}=10, 1, 0.1$ an annular beam structure evolves at a certain time, which depends on τ_{VT}. Inasmuch as the depth of the trough in the intensity profile on the beam axis and its rate of change are determined here by the concentration of the reaction product, which, in turn, depends on τ_{VT}, the longitudinal cross section of its formation (like the length of subsequent focusing of the beam) also depends on τ_{VT}.

A scenario of the time evolution of beam focusing can be conceived by comparing the optical intensity profiles at different times for various relaxation times τ_{VT} and cross sections z. At a certain time following the formation of the region of high concentration of the reaction product, an annular beam structure is formed in a particular cross section of the medium z_0, causing light to be focused in a cross section z_0+z_1. Diffusion causes the transverse width of the high-N region to increase with time for $z<z_0$. As a result, the reaction product acquires a distribution that is nearly uniform within the limits of the beam, so that the beam loses its annular profile and is transformed into a nearly Gaussian shape, and the optical focus at z_0+z_f disappears. Because of time lag, the concentration in the medium lying after the cross section z_0 begins to increase later than in the region lying before z_0, causing the Gaussian profile to transform into an annular profile in a new cross section z_0+z_1, so that now the beam is focused in a new cross section $z_0+z_1+z_{f1}$. Subsequently, after a certain time lapse, a beam structure having a single maximum is also formed already within the limits of the region $z<z_0+z_1$. The focusing process then repeats. Consequently, like the interval of the medium in which the light intensity has a minimum on the axis, the beam focus moves lengthwise along the cell. However, the direction of motion of the focus is no longer monotonic, owing to the variation of the transverse width of the beam as a result of absorption of its light energy and diffraction.

We have already observed these features of the dynamics of the light intensity profile in previous modeling of the propagation of annular beams in a chemically active mixture,[12] including a detailed analysis of the motion of the optical focus. For the case treated in the present paper the beam initially has a Gaussian profile. The laws describing the motion of its focus therefore differ from those described in Ref. 12.

In particular, the focus of the beam probably moves from larger to smaller z in the initial stage of formation of the high-concentration region, and only then does the above-described dynamics of its motion take place.

It follows from these considerations that beam diffraction (like saturation of the absorption coefficient) in a certain time interval can cause the light intensity to increase in part of the medium. In view of the complex dynamics of the transformation of the beam intensity profile, generally speaking, the intensity can undergo several time oscillations on the axis of the initially Gaussian beam.

b) Case $n=1$

We now analyze the effectiveness of controlling a chemical reaction when light energy is absorbed by the first excited molecular vibrational level for the parameters

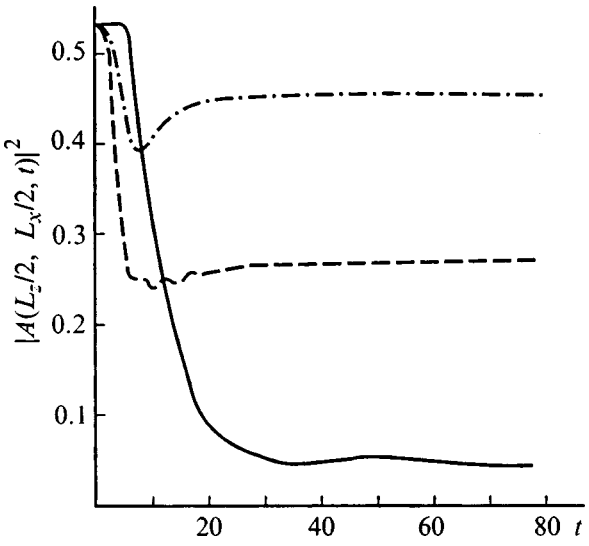

FIG. 5. Time evolution of the light intensity at the center of the cell ($z=L_z/2$, $x=L_x/2$) with the absorption of light energy by the first vibrational transition ($n=1$) for $D=10^{-2}$, $q=50$, $\delta_0=4$, $\sigma=10$, and various relaxation times: $\tau_{VT}=0.1$ (solid curve), 1 (dashed curve), 10 (dot-and-dash curve).

$$n=1, \quad q=50, \quad \sigma=10, \quad 0<t\leqslant 75, \quad \delta_0=4,10. \qquad (4)$$

We note that the time interval up to the onset of explosive absorption is shortened by increasing q and σ. It is important to emphasize that one value of $\delta(E)$ corresponds to two values of E for the given dependence of the absorption coefficient on the vibrational energy over a large range of the latter. Consequently, after the rapid departure of the vibrational energy from its equilibrium value as a result of the onset of explosive absorption, the absorption coefficient can lie either on the rising part $\delta(E)$ (positive feedback) or on the descending part of $\delta(E)$ (negative feedback), depending on the time τ_{VT} and the parameter ε, which characterizes the rate of change of the temperature of the mixture. It is obvious that the stimulation of chemical reaction will follow a different scenario under these conditions.

The main laws discerned for the preceding case are still valid, but they become even more pronounced. However, owing to the fundamentally different dependence of the absorption coefficient on the vibrational energy (it now has a maximum), there are qualitative differences from the case $n=0$. For example, the time dependence of the intensity at the center of the cell (Fig. 5) has a distinct minimum only for $\tau_{VT}=10$. It is important to emphasize that the beam intensity undergoes several oscillations. The period and number of such oscillations depends strongly on the relaxation time of the vibrational energy: The period is equal to 40 for $\tau_{VT}=0.1$, is of the order of 6 for $\tau_{VT}=1$, and is of the order of 3 for $\tau_{VT}=10$. The time of onset of the light intensity oscillations also differs, decreasing as τ_{VT} increases. Oscillations take place both in the stage of the partially formed domain of high absorption ($\tau_{VT}=0.1$) and in the time interval of maximum departure of the vibrational energy from its equilibrium value ($\tau_{VT}=1$), as well as in the explosive absorption stage. All three situations have in common the development of intensity oscillations after formation of the re-

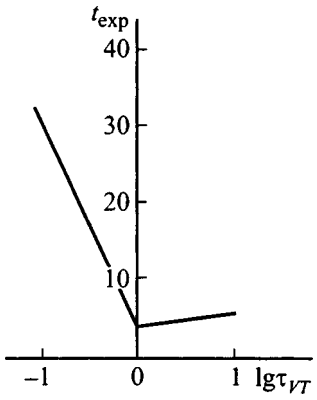

FIG. 6. Dependence of the time of onset of efficient reaction on the logarithm of the relaxation time of vibrational energy at the center of the cell ($z=L_z/2$, $x=L_x/2$) for the following parameters of propagation of diffracted beams, $D=10^{-2}$, $q=50$, $\delta_0=4$, and $\sigma=10$ with the absorption of light energy by the first vibrational transition ($n=1$).

gion of high concentration of the reaction product, wherein the optical density is sufficient for the formation of an annular beam profile. This result evinces the diffractiona l origin of the oscillations. They can be amplified by the existing dependence of the absorption coefficient on the vibrational energy if the absorption varies near the maximum.

The dependence of the time of onset of efficient reaction on τ_{VT} also exhibits qualitative changes (cf. Figs. 2a and 6): we see that, as light energy is absorbed in the vibrational transition $1\rightarrow 2$, $t_{\exp}$ no longer icreases monotonically with τ_{VT} but instead there is an initial decrease in $t_{\exp}$ followed by a slow increase. This complex dependence is attributable to the increase in the absorption coefficient as the product concentration and the vibrational energy increase in the initial stage of interaction and to its decrease with a further increase in E. We note that the substantial increase in the relaxation time of efficient reaction for small values of τ_{VT} is due to the essentially quadratic dependence of the absorption coefficient on a small quantity (N and E are small in the initial stage of interaction). In the interval of times $\tau_{VT}=1-10$ the behavior of $t_{\exp}$ is similar to the case of absorption of light energy by the ground vibrational transition.

The initial part of the dependence of $t_{\exp}$ on τ_{VT} can probably be explained as follows. For small τ_{VT} the rapid temperature increase near the input cross section of the cell is accompanied by the formation of a domain of high absorption, which then begins to move along the cell. Since the absorption coefficient is proportional to E in the case analyzed here, the velocity of the domain is also proportional to E. For small τ_{VT} the nonuniformity in the longitudinal distribution of the vibrational energy is clearly pronounced by virtue of the strong absorption of light energy by the newly formed domain. As τ_{VT} increases, the formation of the domain near the front wall of the cell slows down, the absorption coefficient becomes more uniform along the medium (within the boundaries of the incipient domain), and the domain evolves more uniformly with time. In this case, accordingly, the value of $t_{\exp}$ for $\tau_{VT}\approx 0.1$ represents the arrival time of the leading edge of the domain of high absorption at the cross section $z=L_z/2$.

The foregoing conclusion is also confirmed by an analysis of the dynamics of formation of the longitudinal domain of concentration of the reaction product. The steady-state distribution of N along z for $\delta_0=10$ is shown in Fig. 3b. Two observations are relevant in this regard. First, for $\delta_0=10$ and $\tau_{VT}=0.1$ the boundary of the domain is situated to the left of the cross section $z=0.5$ (Fig. 4 shows the results of the calculations for $\delta_0=4$). Second, our analysis of the case of weak heat removal has important bearing on the established spatial distributions. In the presence of fully developed thermal diffusion the structure of the absorption domain can undergo a more pronounced change for $\tau_{VT}\geqslant\tau_D$ (τ_D is a time constant characterizing the thermal conductivity of the medium).

In contrast with the preceding case, the transverse structure of the domain of high absorption is also transformed. For example, for small relaxation times (say, $\tau_{VT}=0.1$) the diffraction of radiation essentially fails to broaden the high-N region with increasing z. The structure of the level lines of the reaction product concentration is similar to the case of weakly diffracted beams.

When τ_{VT} is increased to 1, diffraction of the beam influences the region of efficient reaction only in part of the cell. For large relaxation times the transverse cross section of the domain of high concentration increases along z, repeating the broadening of the beam radius.

An analysis of the light intensity profile shows that the motion of the focus retraces the path encountered in the absorption of light energy by the ground transition. The only difference is more distinct focusing of the beam both in terms of its peak intensity and in terms of its radius (it decreases by a factor of 1.5–2 at the exit from the medium for the parameters compared). This feature indicates sharper transverse boundaries of the domain of high absorption.

c) Case $n=2$

The bistable dependence $E(T)$ mentioned in Ref. 3 for a chemically neutral medium can also occur in this case. Consequently, the unambiguous interpretation of the results of the numerical simulations requires a preliminary investigation of critical points of the system of equations describing the stimulation of chemical reaction with the injection of light energy through the $2\rightarrow 3$ transition within the framework of a point model. This problem is of independent interest and is beyond the scope of the present study. Here we stress the fact that the laws governing the laser-induced chemical kinetics of a gas reaction with the absorption of light energy by the vibrational transition $2\rightarrow 3$ ($n=2$) for $T_0=0.25$ are similar to those obtained for $n=1$ and $T_0=0.125$.

CONCLUSIONS

It is evident from the foregoing discussion that allowance for processes involving the relaxation of absorbed light energy into heat in the laser stimulation of reversible chemical reactions fundamentally alters their kinetics from the laws obtained on the basis of instantaneous relaxation. In particular, the stimulation process has a distinctly time-

dependent character due to bleaching of the medium as a result of saturation of the vibrational transition.

The complex dependence of the absorption coefficient on the vibrational energy and beam diffraction can lead to the development of oscillations (damped in the given situation) of the concentration of the reaction product in part of volume of the medium.

The influence of the diffraction of optical radiation having a Gaussian intensity profile at the entrance to the cell results in focusing of the light beam due to the formation of a trough in the intensity distribution along the axis of the beam as it propagates. Saturation of the absorption coefficient, on the other hand, causes the focus of the beam to move.

This work has received partial support form the Russian Fund for Fundamental Research (Grant No. 95-02-04448).

[1] G. A. Kotel'nikov and A. N. Nefedov, At. Tekh. za Rubezh., No. 3, 8 (1986).
[2] A. M. Bonch-Bruevich and M. N. Libenson, Izv. Akad. Nauk SSSR, Ser. Fiz. **46**, 1104 (1982).
[3] A. N. Oraevskiĭ and I. E. Protsenko, Tr. Fiz. Inst. Ross. Akad. Nauk **187**, 184 (1988).
[4] A. Nitzan and J. Ross, J. Chem. Phys. **59**, 241 (1973).
[5] F. V. Bunkin, N. A. Kirichenko, and B. S. Luk'yanchuk, Izv. Akad. Nauk SSSR, Ser. Fiz. **46**, 1150 (1982).
[6] E. A. Gel'man, A. D. Margolin, and V. M. Shmelev, Khim. Fiz. **9**, 1150 (1990).
[7] O. A. Gunaze and V. A. Trofimov, Bull. Russian Acad. Sci. Ser. Phys. (Suppl. Physics of Vibrations) **57**(4), 193 (1993).
[8] O. A. Gunaze and V. A. Trofimov, Izv. Akad. Nauk SSSR, Ser. Fiz. **60**, No. 6 (1996).
[9] O. A. Gunaze, A. G. Gushchina, and V. A. Trofimov, Khim. Fiz. **13**(11), 24 (1994).
[10] V. S. Letokhov and V. P. Chebotayev, *Nonlinear High-Resolution Laser Spectroscopy* [Springer-Verlag, Berlin–New York, 1977; Nauka, Moscow, 1990, 510 pp.].
[11] V. M. Akulin and N. V. Karlov, *Intense Resonance Interactions in Quantum Electronics* [Springer-Verlag, Berlin–New York, 1992; Nauka, Moscow, 1987, 312 pp.].
[12] O. A. Gunaze and V. A. Trofimov, Zh. Tekh. Fiz. **68**(3), 8 (1998) [Tech. Phys. **43**, 271 (1998)].

Translated by James S. Wood

Total internal reflection of a Gaussian light beam

P. D. Kukharchik, V. M. Serdyuk, and I. A. Titovitskiĭ

Belarus State University, A. N. Sevchenko Scientific-Research Institute of Applied Physics Problems, 220064 Minsk, Belarus

(Submitted December 20, 1996)
Zh. Tekh. Fiz. **69**, 74–78 (April 1999)

An approximate theoretical model for calculating the reflected and refracted fields of a Gaussian light beam at a plane interface between two isotropic media is formulated on the basis of a Fourier integral. In the vicinity of the critical angle of incidence (for total internal reflection) the model predicts the presence of two refracted beams, one displaced along the interface by an amount equal to the Goos–Hänchen shift; curvature of the phase fronts and nonalignment of the effective directions of energy and phase propagation occur for each beam, as in an anisotropic medium. © *1999 American Institute of Physics.* [S1063-7842(99)01304-5]

Wave reflection and refraction phenomena are often encountered in the most diverse branches of physics. One would think that these phenomena could be described by a simple and reliable model based on the plane-wave approximation and the Fresnel equations.[1–4] However, such an approximation is not entirely correct in the presence of total internal reflection. In reality, in the vicinity of a plane interface between two transparent media the field in a refracting medium is described by a homogeneous traveling wave for angles of wave incidence arbitrarily smaller than the angle of total internal reflection (the critical angle) and is described by an inhomogeneous wave decaying in the direction away from the boundary for angles of incidence arbitrarily greater than the critical angle.[1–4] The model system experiences an abrupt change in its qualitative state, but such a change does not really take place.[2,3] The problem is that the actual wave fields are always bounded and inhomogeneous in space, and the reflection and refraction laws for spatially confined wave beams differ from the laws for an infinite plane wave.[2–4] The disparity is especially significant in the presence of total internal reflection.

So far theoretical studies of the total internal reflection of spatially confined beams have been reduced to investigations of the reflected field, because excessively cumbersome integral equations are obtained for the refracted field.[2–6] To the best of our knowledge, there is only one paper[7] in which the refracted field of an inhomogeneous light beam subjected to total internal reflection has been investigated theoretically.

In that paper an exact solution of the differential equation for the wave function of the refracted field has been constructed in the form of a double integral, and a simple asymptotic approximation has been found for it, but certain effects such as, for example, the Goos–Hänchen shift[3,5,6] cannot be described in this approximation. The objective of the present study is to develop a more realistic approximation of the theory of field calculations in the reflection and refraction of a Gaussian beam, which is the simplest model of a spatially confined light field.

Let a Gaussian light beam be incident from the negative x direction at an angle other than $\pi/2$ on a plane interface between two isotropic media at $x=0$ with dielectric permittivities ε_i (at $x<0$) and ε (at $x>0$), and let the wave field of the beam at the interface be written in the form

$$U_i(0,z)=\exp(-z^2/w^2+i\beta_b z), \tag{1}$$

where β_b and w are the propagation constant and the effective half-width of the beam along the interface [the time factor $\exp(-i\omega t)$ is omitted here, and the dependence of the field on the tangential coordinate y, which is perpendicular to the plane of incidence, is disregarded).

The dielectric permittivities ε_i and ε can be complex, i.e., absorption is admitted in both media. We expand the function (1) into a Fourier integral in plane-wave modes with propagation constants β along the interface, whereupon the field of the incident beam near the interface is represented by the expression[3,4]

$$U_i(x,z)=\int_{-\infty}^{+\infty} A(\beta)\exp(i\alpha_i x+i\beta z)\,d\beta;$$

$$A(\beta)=(w/2\pi^{1/2})\exp(-w^2[\beta-\beta_b]^2/4), \tag{2}$$

where $\alpha_i=(k_0^2\varepsilon_i-\beta^2)^{1/2}$ is the component, normal to the interface, of the wave vector of each (ith) Fourier component of the field, and $k_0=\omega/c$ is the wave number.

We also seek the reflected field U_r and the refracted field U in Fourier integral form. Allowing for the fact that each plane-wave Fourier component is reflected and refracted independently of the others, we can at once write integral equations for these fields:[3,4]

$$U_r(x,z)=\int_{-\infty}^{+\infty} R(\beta)A(\beta)\exp(-i\alpha_i x+i\beta z)\,d\beta;$$

$$U(x,z)=\int_{-\infty}^{+\infty} T(\beta)A(\beta)\exp(i\alpha x+i\beta z)\,d\beta, \tag{3}$$

where

$$\alpha=(k_0^2\varepsilon-\beta^2)^{1/2};\quad R(\beta)=(\alpha_i\Theta-\alpha\Theta_i)/(\alpha_i\Theta+\alpha\Theta_i),$$

$$T(\beta)=1+R(\beta)=2\alpha_i\Theta/(\alpha_i\Theta+\alpha\Theta_i) \tag{4}$$

are the Fresnel reflection and refraction coefficients;[1–4] $\Theta = \Theta_i = 1$ for TE polarization of the incident light field, and $\Theta = \varepsilon$, $\Theta_i = \varepsilon_i$ for TM polarization.

It will be helpful in the ensuing discussion to represent the coefficients R and T (4) in terms of the parameter v, which we call the reflection parameter:

$$R(\beta) = \exp(-2v); \quad T(\beta) = 1 + \exp(-2v);$$

$$v = \tanh^{-1}(\alpha_i \Theta / \alpha \Theta_i) = \frac{1}{2} \ln\left(\frac{1 + \alpha \Theta_i / \alpha_i \Theta}{1 - \alpha \Theta_i / \alpha_i \Theta} \right). \tag{5}$$

The physical significance of this parameter is totally apparent; its real and imaginary parts characterize the variations of the intensity and the phase of a plane wave in reflection.

To find the spatial distribution of the reflected and refracted fields, it is sufficient to substitute Eqs. (4) and (5) into (3) and to evaluate the resulting integrals. Integral equations of this kind are often met in various problems involving the propagation of wave beams in layered systems,[3,8] in particular, for calculations of exciting fields in planar dielectric waveguides.[9] In this article, therefore, together with the specific results of calculations we also note some unconventional methods for performing such calculations.

In calculating integrals of this kind, it is customary to use an approximation based on the assumption that the light wavelength $\lambda = 2\pi/k_0$ is small in comparison with the half-width of the beam w, i.e., that the condition $k_0 w \gg 1$ holds.[3–6,8] The function $A(\beta)$ in the integrals (2) and (3) then differ appreciably from zero in a very narrow interval of the variable of integration $\beta_b(1 - 2/\beta_b w) < \beta < \beta_b(1 + 2/\beta_b w)$, so that the linear approximation in the argument β can be used for the parameter α_i:

$$\alpha_i(\beta) \approx \alpha_{ib} - \sigma_i(\beta - \beta_i); \quad \alpha_{ib} = (k_0^2 \varepsilon_i - \beta_b^2)^{1/2};$$

$$\sigma_i = \beta_b \mathrm{Re}(1/\alpha_{ib}). \tag{6}$$

The normal component of the averaged wave vector of the beam α_{ib} can be complex-valued by virtue of absorption, but the smallness of the refractive index means that the imaginary part of the parameter σ_i can be ignored. In the opposite case, i.e., for strong absorption in a semi-infinite medium, the exciting field (2) becomes negligible at the interface $x = 0$.

Substituting the expansion (6) into Eq. (2), for the incident field we obtain

$$U_i(x,z) = \exp(-[z - \sigma_i x]^2 / w^2 + i[\beta_b z + \alpha_{ib} x]).$$

This equation is the well-known expression for the field of a Gaussian beam, neglecting transverse diffractional spreading of the amplitude (in the geometrical-optics approximation), and it is valid at small distances from the interface in question ($|x| \ll k_0 w^2$).[4,8] Here the phase of the field depends on the longitudinal coordinate of beam propagation ($\beta_b z + \alpha_{ib} x$), and the intensity depends on the transverse coordinate ($z - \sigma_i x$).

For the reflected and refracted fields this sort of simple procedure of linearizing the spectral dependence of α on β works only for beams at angles of incidence much smaller or much greater than the critical angle, i.e., when $\beta_b(1 + 4/\beta_b w) < \beta_c$ or $\beta_b(1 - 4/\beta_b w) > \beta_c$, where $\beta_c = k_0(\mathrm{Re}\,\varepsilon)^{1/2}$. The following approximation is then valid:

$$\alpha(\beta) \approx \alpha_b - \beta_b(\beta - \beta_b)/\alpha_b; \quad \alpha_b = (k_0^2 \varepsilon - \beta_b^2)^{1/2}. \tag{7}$$

However, if β_c lies in the interval of integration $\beta_b(1 - 2/\beta_b w) \leqslant \beta \leqslant \beta_b(1 + 2/\beta_b w)$, the parameter α_b is found to be very small or to vanish altogether, rendering the approximation (7) invalid. To circumvent this difficulty, we expand the parameter α with respect to β, not at the point β_b, but at the point $\beta = \beta_b + s \equiv \beta_w$, where s is a certain complex quantity. Calculations for various values of this quantity and a comparison of the results with exact calculations[7] show that the optimum value is $s = -(1 + i)/2w$, which corresponds to the expansion

$$\alpha(\beta) \approx \alpha_w - \beta_w(\beta - \beta_w)/\alpha_w \approx \alpha_B - \sigma(\beta - \beta_b), \tag{8}$$

where

$$\beta_w = \beta_b - (1 + i)/2w;$$

$$\alpha_w = (k_0^2 \varepsilon - \beta_w^2)^{1/2} \approx [k_0^2 \varepsilon - \beta_b^2 + \beta_b(1 + i)/w]^{1/2};$$

$$\alpha_B = \alpha_w - \beta_b(1 + i)/2w\alpha_w; \quad \sigma = \beta_b / \alpha_w. \tag{9}$$

The first term of the expansion (8) α_B differs from its counterpart α_b in the expansion (7) by the added complex quantity in the square root. Obviously, for large values of the difference $|k_0^2 \varepsilon - \beta_b^2|$, i.e., far from the region of total internal reflection, this increment can be disregarded, but near the critical angle of incidence it renders the coefficient σ (9) finite (at least in a nonamplifying medium, with nonnegative $\mathrm{Im}\, e$) and thus ensures that the linear representation (8) will be valid.

Equations (6) and (8) can be used to find an analogous approximation for the reflection parameter (5):

$$v \approx v_B + q(\beta - \beta_b); \quad v_B = \tanh^{-1}\left(\frac{\Theta_i \alpha_B}{\Theta \alpha_{ib}} \right);$$

$$q \approx \frac{\Theta_i \Theta k_0^2 (\varepsilon - \varepsilon_i) \sigma}{\alpha_{ib}(\Theta^2 \alpha_{ib}^2 - \Theta_i^2 \alpha_b^2)} \tag{10}$$

or, since σ is not very large except at near-critical angles of incidence,

$$q \approx -\Theta_i \sigma / \Theta \alpha_{ib}.$$

Substituting Eqs. (5), (6), (8), and (10) into Eq. (3), we can calculate the spatial distribution functions of the reflected and refracted fields. For the reflected field we obtain

$$U_r(x,z) = \exp(-2v_B + 4q_R^2 w^{-2} - [z + \sigma_i x - 2q_I]^2 w^{-2} + i\{\beta_b z - \alpha_{ib} x - 4q_R[z + \sigma_i x - 2q]\}). \tag{11}$$

From now on the subscripts R and I are abbreviated notations for the real and imaginary parts of the quantity to which they are attached.

It follows from Eq. (11) that the effective reflection parameter v_{eff} is given by the equation

$$v_{\mathrm{eff}} = v_B - 2q_R^2 / w^2. \tag{12}$$

Even if allowance is made for the fact that the quantity v_B (10) does not coincide with the value of the reflection parameter $v_F=\tanh^{-1}(\Theta_i\alpha_b/\Theta\alpha_{ib})$ corresponding to the Fresnel equations,[1–4] the real part of v_{eff} is smaller than v_F, so that the average reflection coefficient $R=\exp(-2v_{\text{eff}})$ is somewhat larger than the Fresnel coefficient. Moreover, the beam shifts as a unit whole along the interface (along the z axis) by the amount

$$\Delta z=2q_I\approx 2\Theta_i\beta_b\alpha_{wI}/\Theta k_0(\alpha_{wR}^2+\alpha_{wI}^2)(\varepsilon_i-\varepsilon_R)^{1/2}, \tag{13}$$

i.e., we have the Goos–Hänchen effect,[3–6] which is not identical for TE and TM field polarizations.

Equation (11) describes another phenomenon: The direction of propagation of the phase of the reflected beam, being specified by the imaginary exponent of the exponential function (11), does not coincide with the longitudinal coordinate $(\beta_b z-\alpha_{ib}x)$, which is parallel to the lines of constant intensity of the field. This longitudinal coordinate obviously determines the direction of propagation of the beam energy (11). Consequently, the effective wave normal deviates from the ray normal in the direction of the interface by the angle

$$\delta_r=4\Theta_i\varepsilon_R^{1/2}\alpha_{wR}/\Theta(\varepsilon_i-\varepsilon_R)k_0w^2(\alpha_{wR}^2+\alpha_{wR}^2),$$

which differs for TE and TM field polarizations. For the reflected beam in an isotropic medium, therefore, the induction of anisotropy takes place, eliciting a slight birefringence.

All these effects resulting from the spatial inhomogeneity of the incident field are the most conspicuous at near-critical angles of incidence. We now estimate their order of magnitude in the optimal case:

$$v_{\text{eff}}=v_B-\Theta_i^2\varepsilon_R^{1/2}/\Theta^2(\varepsilon_i-\varepsilon_R)k_0w;$$

$$\Delta z/w=(\Theta_i/\Theta)\varepsilon_R^{1/4}\{2/(\varepsilon_i-\varepsilon_R)k_0w\}^{1/2};$$

$$\delta_r=\{2^{3/2}\Theta_i\varepsilon_R^{1/4}/\Theta(\varepsilon_i-\varepsilon_R)\}(k_0w)^{-3/2}. \tag{14}$$

We compare the last quantity with the aperture angle in which a Gaussian beam must propagate, i.e., with the angle of its diffractional spreading, which has been disregarded above. This angle is equal to $(2/k_0w)$ (Refs. 4 and 8) and becomes much greater than the angle δ_r (14). Hence it follows that the diffractional spreading effect is stronger than the effect of angular shifting of the wave normal and certainly prevents it from being observable experimentally. For the refracted field (3) the imaginary part of the coefficient σ (9) of the linear expansion of α with respect to β (9) cannot be disregarded as it was for the incident and reflected beams in regard to α_i (6). The integrals (3) in the expression for U are now integrals of the form

$$\int_{-\infty}^{+\infty}\exp(-[\beta-\beta_b-iC]^2-2i\sigma x[\beta-\beta_b]/w)d\beta, \tag{15}$$

in which the linear approximation fails for large x; the integrand becomes far too large for $x(\beta-\beta_b)\sigma_I>0$. To surmount this difficulty, we replace the argument $(\beta-\beta_b)$ of the term containing $x\sigma_I$ in Eq. (15) by its average value $iC/2$ over the effective interval of integration. We then have the approximate representation of the refracted field (3)

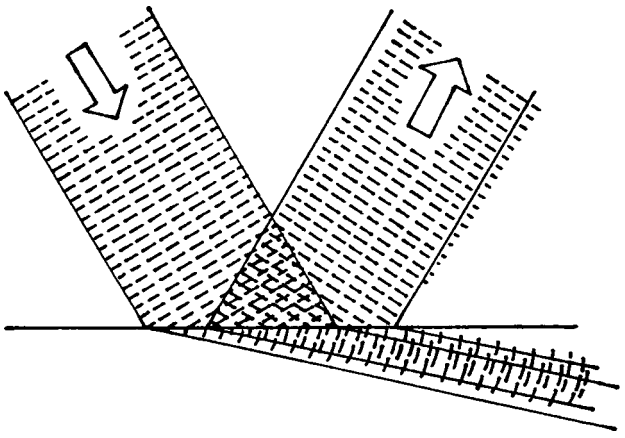

FIG. 1. Diagram of the reflection and refraction of a beam near the critical angle of incidence.

$$\begin{aligned}U(x,z)=\exp(i\alpha_Bx+i\beta_bz)\{&\exp(-[z-\sigma_Rx]^2/\\&w^2+i\sigma_I[z-\sigma_Rx]x/w^2)+\exp(-2v_B+4q_R^2w^{-2}\\&-[z-\sigma_Rx-2q_I]^2/w^2+i[\sigma_Ix-4q_R]\\&\times[z-\sigma_Rx-2q_I]/w^2-2q_R\sigma_Ix/w^2)\},\end{aligned} \tag{16}$$

which is written as the superposition of the fields of two beams. At the interface $x=0$ their amplitudes are equal to the amplitudes of the incident beam (1) and the reflected beam (11), respectively; the second beam is displaced along the interface relative to the first by the Goos–Hänchen shift (13) (Fig. 1). The constant-phas e surfaces of both beams (16), generally speaking, are not planar, and they have a constant curvature $K=2\sigma_R\sigma_I/\beta_bw^2$. The extremum $K_e=-1/2w$ is attained at near-critical angles of incidence. Other effects are maximized in this same region: The reflection parameter for the second refracted beam decreases by the same amount (12) as for the reflected field, and the amplitude attenuation coefficients for the refracted beams differ by the amount $2q_R\sigma_I/w^2$. Another phenomenon, analogous to the induction of anisotropy for the reflected field, is readily discerned: The directions of propagation of the intensity perpendicular to the planes $z-\sigma_Rx=\text{const}$ and $z-\sigma_Rx-2q_R=\text{const}$ do not coincide with the directions of propagation of the phase, which are governed mainly by the coordinate dependence $(\alpha_Bx+\beta_bz)$ in Eq. (16). It is characterized by the angle

$$\delta\approx(1-\alpha_{BR}\sigma_R/\beta_b)(1+\sigma_R^2)^{-1/2}$$

between the effective phase vector and the ray normal, whose maximum value $\delta\approx(\beta_bw)^{-1/2}$ is again attained in the region of near-critical angles of incidence.

If we make use of the fact that wave beams used in optics and acoustics usually obey the condition $(k_0w)^{1/2}\gg 1$, the expressions for the reflected and refracted fields (11) and (16) are simplified:

$$U_r(x,z)\approx R_B\exp(i[\beta_bz-\alpha_{ib}x]-[z+\sigma_ix]^2/w^2);$$

$$\begin{aligned}U(x,z)\approx T_B\exp(i[\beta_bz+\alpha_Bx]-[z-\sigma_Rx]^2/\\w^2+i\sigma_I[z-\sigma_Rx]x/w^2),\end{aligned} \tag{17}$$

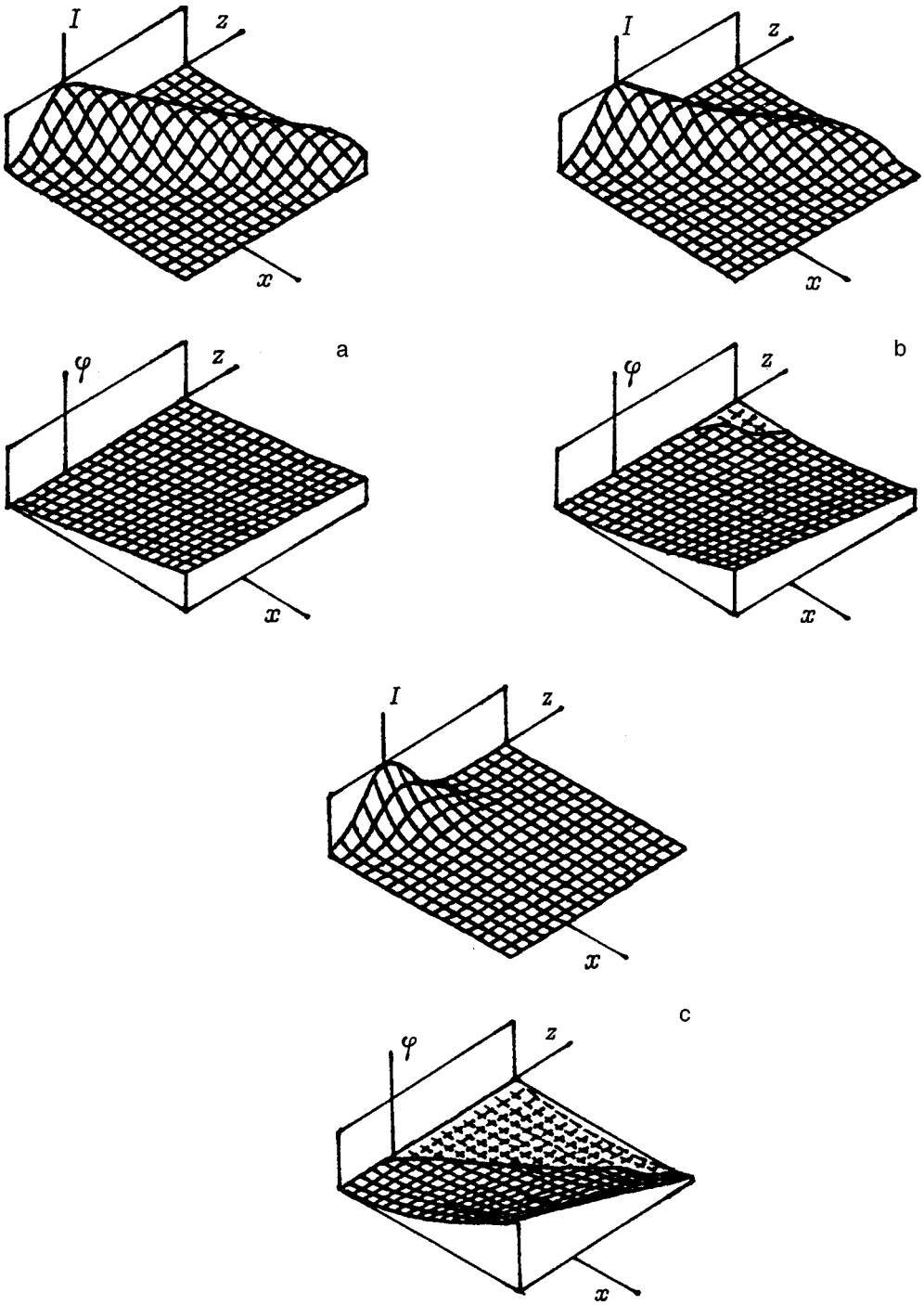

FIG. 2. Spatial distributions of the intensity I and the phase φ of the refracted at the interface of two transparent media with refractive indices $\varepsilon_i=1.6$ and $\varepsilon^{1/2}=1.5$ at angles of incidence 5″ smaller than the critical angle (a), equal to the critical angle (b), and 5″ larger than the critical angle (c). The scale along the x axis is magnified 200× relative to the scale along the z axis.

where $R_B=\exp(-2v_B)=(\alpha_{ib}\Theta-\alpha_B\Theta_i)/(\alpha_{ib}\Theta+\alpha_B\Theta_i)$ and $T_B=1+R_B=2\alpha_{ib}\Theta(\alpha_{ib}\Theta+\alpha_B\Theta_i)$ are the reflection and transmission coefficients for the averaged directions of propagation.

In application to such beams almost all the above-analyzed phenomena are found to be negligible, probably with the exception of curvature of the wave front and the difference in the directions of the phase and ray normal for the refracted beam.

The graphs in Fig. 2 illustrate the spatial structure of the field (17) refracted at the interface of two transparent media when $k_0w=4.1\times10^4$. The rapid change of phase along the z axis as specified by the factor $\exp(i\beta_bz)$ has been ignored in plotting the graphs. A comparison of Fig. 2 with corresponding graphs in Ref. 7 shows that the approximation (17) far better approximates the exact solution than the approximate solution obtained in Ref. 7. In particular, it can be used to describe the nonlinear dependence of the phase on the normal coordinate x. Figure 2 and Eqs. (9) and (17) show that the attenuation mechanism in a refracting medium comes

into play gradually until the increasing refraction angle attains the critical value (approximately for $\beta_c-\beta_b\leqslant 1/8w$ and $\beta_c=k_0\varepsilon_R^{1/2}$), and the oscillatory dependence on the normal coordinate vanishes a little later (for $\beta_b-\beta_c\geqslant 1/8w$). This process takes place as though the spatial inhomogeneity of the beam somewhat exceeded the refractive index and induced a small absorption index in the transparent medium [see the expression for α_w (9)]. The presence of intrinsic absorption in a refracting medium makes these beam confinement effects less pronounced,[7] and in an amplifying medium ($\mathrm{Im}\,\varepsilon<0$) the above-developed theory is valid only for $|\mathrm{Im}\,\varepsilon|<1/k_0w$, i.e., for very low gains.

[1] M. Born and E. Wolf, *Principles of Optics*, 4th ed. [Pergamon Press, New York, 1970; Nauka, Moscow, 1973, 720 pp.].
[2] V. A. Kizel', *Reflection of Light* [in Russian], Nauka, Moscow, 1973, 352 pp.
[3] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media*, Vols. 1 and 2 [Springer-Verlag, Berlin–New York, 1990, 1992; Nauka, Moscow (1989), 414 pp.].
[4] M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *Wave Theory* [in Russian], Nauka, Moscow, 1979, 224 pp.
[5] F. Goos and H. Hänchen, Ann. Phys. (Leipzig) **1**, 333 (1947).
[6] V. Shah and T. Tamir, J. Opt. Soc. Am. **73**, 37 (1983).
[7] V. M. Serdyuk, Dep. VINITI No. 3092-V92, All-Union Institute of Scientific and Technical Information, Moscow (1992).
[8] A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* [Wiley, New York, 1984; Mir, Moscow, 1987, 616 pp.].
[9] A. M. Bel'skiĭ, V. V. Mogil'nyĭ, and V. B. Movchan, Opt. Spektrosk. **65**, 1308 (1988) [Opt. Spectrosc. **65**, 773 (1988)].

Translated by James S. Wood

Electromagnetic pulse amplification in a Cherenkov laser

S. Oganesyan

''Lazernaya Tekhnika'' Research and Development Company, 375090 Yerevan, Armenia
(Submitted July 30, 1997)
Zh. Tekh. Fiz. **69**, 79–83 (April 1999)

An investigation is made of the amplification of a Gaussian electromagnetic pulse in a Cherenkov waveguide laser for the cases of long and short waveguides. It is shown that in the first case, the concept of a characteristic pulse duration τ_0 can be introduced. It is established that when the pulse duration is short ($\tau<\tau_0$) the gain is determined only by its spectral width, and the amplification process leads to a change in the pulse envelope. It is shown that in a short waveguide pulse amplification can be achieved without any change in shape.
 [S1063-7842(99)01404-X]

INTRODUCTION

The amplification process in a Cherenkov waveguide laser has been studied by many authors[1–7] in cases where an external monochromatic probe wave is supplied to the system (note that the present author[8] has considered the case where the amplified radiation is formed from the spontaneous noise). Various waveguide shapes have been analyzed in the literature to enhance the efficiency of interaction between the electron beam and the surface wave. These range from the simplest planar and cylindrical waveguides to more complex configurations.[4–6] The amplification mechanism in a Cherenkov waveguide laser depends on the quality of the electron beam and the waveguide size.[1] Walsh[2] showed that if the electron density ρ_0 is high (the high-gain case), and if the electron beam spread is small and the waveguide is fairly long, instability may appear in the waveguide-mode–electron-beam system with the growth rate $G\sim\sqrt[3]{\rho_0}$. If the electron density is low (the low-gain case), the gain is $G\sim\rho_0$ (Ref. 1). In this limit the gain is determined by the ratio between the spread of the z projection of the velocity Δv_z and the spread of the z projection of the photon wave vector $\Delta k_z=2\pi/L$ due to the finite waveguide length L. For a long waveguide ($\Delta v_z/v_z>\Delta k_z/2\pi k_z$) the simplest amplification mechanism is based on the imposition of a static magnetic field along the electron beam.[1] For a short ($\Delta v_z/v_z<\Delta k_z/2\pi k_z$) waveguide (Ref. 3) the magnetic field plays a focusing role and the amplification mechanism is based on the fact that the momentum conservation law is not satisfied along the waveguide.[1] It was noted that all these results were obtained assuming that the amplified signal is monochromatic. In practice, however, the probe wave always has a finite duration τ. In the present paper a detailed analysis is made of the electromagnetic pulse gain as a function of its width $\Delta\omega=1/\tau$ in long and short waveguides. It is shown that in the first case, the gain depends strongly on the pulse duration τ and is accompanied by a change in shape. In the second case, we find the condition under which amplification takes place without any change in shape.

PROPAGATION OF AN ELECTROMAGNETIC PULSE IN A PLANAR WAVEGUIDE

Let us assume that a monochromatic TM wave propagates along the z axis in a planar waveguide. We shall assume that the symmetry plane of the waveguide coincides with the yz plane and its thickness is $2a$. The projections of the electric and magnetic fields are then given by

$$
\begin{aligned}
&|x|>a,\\
&E_z=E_{1z}\exp[i(k_z z-\omega t)\mp q_x x],\ \ E_y=0,\ \ E_x=\frac{ik_z}{q_x}E_z,\\
&H_z=0,\ \ H_y=\frac{i\omega}{q_x c}E_z,\ \ H_x=0;\qquad (1)\\
&|x|<a,\\
&E_x=E_{2z}\sin(k_{2x}x)\exp[i(k_z z-\omega t)],\ \ E_y=0,\\
&E_x=\frac{ik_z}{k_{2x}}E_{2z}\cos(k_{2x}x)\exp[i(k_z z-\omega t)],\\
&H_x=0,\ \ H_y=\frac{i\varepsilon\omega}{ck_{2x}}E_{2z}\cos(k_{2x}x)\exp[i(k_z z-\omega t)],\\
&H_z=0.\qquad (2)
\end{aligned}
$$

Here ε is the dielectric constant of the waveguide. The wave amplitudes E_{1z} and E_{2z} are related by

$$E_{2z}=E_{1z}\frac{\exp(-q_x a)}{\sin(k_{2x}a)}. \qquad (3)$$

Without loss of generality, we can assume that E_{1z} is a real quantity. The frequency and projection of the wave vector of the natural waveguide modes are related by

$$\tan(k_{2x}a)=\varepsilon q_x/k_{2x}, \qquad (4)$$

where

$$q_x=\left(k_z^2-\frac{\omega^2}{c^2}\right)^{1/2},\quad k_{2x}=\left(\varepsilon\frac{\omega^2}{c^2}-k_z^2\right)^{1/2}. \qquad (5)$$

We will be interested below in millimeter electromagnetic waves propagating synchronously ($k_z=\omega/v_z$) or almost synchronously ($k_z\approx\omega/v_z$) with the electron beam. It follows from Eq. (4) that each velocity v_z corresponds to a set of natural waveguide modes having the frequencies $\omega_i(v_z)$ ($i=0,1,\ldots$). The entire following analysis refers only to a single waveguide mode.

Let us now assume that an electromagnetic pulse of duration τ propagates in the same waveguide. Its field may be represented as a superposition of the fields (1) and (2)

$$\mathbf{E}=\int d\omega\,\mathbf{E}(\omega)\exp[i(k_z z-\omega t)-q_x x]\quad\text{and so on.}\qquad(6)$$

For conciseness we shall only consider the z projection of the electric field of the wave. To be specific, we shall assume that the spectral expansion of the field has a Gaussian profile

$$E_z(\omega)=\frac{\tau}{2\sqrt{\pi}}E_{1z}\exp\left[-\frac{(\omega-\omega^0)^2\tau^2}{4}\right]\qquad(7)$$

and we shall find the pulse envelope in the laboratory coordinate system. It is clearly impossible to integrate Eq. (6) exactly. In order to simplify the problem we shall assume that the pulse spectral width $\Delta\omega=1/\tau$ is not very large. We shall also assume that in this frequency range the dispersion of the waveguide material can be neglected: $\varepsilon(\omega)=\varepsilon=\text{const}$ (this is easily satisfied in a quartz crystal, for example). Suppose that ω^0, q_x^0, k_{2x}^0, and k_z^0 are the frequency and projections of the wave vector of the carrier wave. It follows from Eqs. (4) and (5) that for frequencies ω near ω_0 the projections of the wave vector $\mathbf{k}$ are given by

$$k_z=k_z^0+\nu_1\delta\omega,\quad q_x=q_x^0+\nu_2\delta\omega,\quad k_{2x}=k_{2x}^0+\nu_3\delta\omega.\qquad(8)$$

Here the detuning is $\delta\omega=\omega-\omega^0$ (only linear terms with respect to this parameter were taken into account in the calculations),

$$\nu_1=\frac{\varepsilon\omega^0}{c^2k_z^0}\,\frac{q_x^0a[(k_{2x}^0)^2+\varepsilon^2(q_x^0)^2]+2[(k_{2x}^0)^2+\varepsilon(q_x^0)^2]}{q_x^0a[(k_{2x}^0)^2+\varepsilon^2(q_x^0)^2]+2\varepsilon[(k_{2x}^0)^2+(q_x^0)^2]},$$

$$\nu_2=(\nu_1k_z^0-\omega^0/c^2)/q_x^0,\quad \nu_3=(\omega^0\varepsilon/c^2-\nu_1k_z^0)/k_{2x}^0.\qquad(9)$$

The formulas (8) hold when

$$\delta\omega<\min\{k_z^0\nu_1^{-1},\ q_x^0\nu_2^{-1},\ k_{2x}^0\nu_3^{-1}\}.$$

Substituting Eq. (8) into Eq. (6), we find that the z projections of the electric field outside the waveguide ($x>a$ corresponds to $+$ and $x<-a$ to $-$) and inside the waveguide ($-a<x<a$) are respectively given by

$$E_z=\tilde{E}_{1z}^{\pm}(\eta_1^{\pm})\exp(i\Phi_1^{\pm}),\qquad(10)$$

$$E_z=E_{2z}^{+}(\eta_2^{+})\exp(i\Phi_2^{+})+E_{2z}^{-}(\eta_2^{-})\exp(i\Phi_2^{-}).\qquad(11)$$

Here the following notation is introduced:

$$\Phi_1^{\pm}=k_z^0z-\omega^0t\pm iq_x^0x,\quad \eta_1^{\pm}=\nu_1z\pm i\nu_2x-t,$$

$$\Phi_2^{\pm}=k_z^0z-\omega^0t\pm k_{2x}^0x,\quad \eta_2^{\pm}=\nu_1z\pm\nu_3x-t.\qquad(12)$$

The amplitudes $\tilde{E}_{1z}^{\pm}$ and $E_{2z}^{\pm}$ have the simplest form when the terms ν_2x and ν_3x in formulas (12) are small ($|\nu_2x|\ll\tau$, $|\nu_3x|\ll\tau$),

$$\tilde{E}_{1z}^{\pm}=\tilde{E}_{1z}=E_{1z}\exp\left[-\frac{(\nu_1z-t)^2}{\tau^2}\right],$$

$$E_{2z}^{\pm}=\mp\frac{i}{2}\,\tilde{E}_{2z}\mp\frac{1}{2}k_{2x}^0a\left(\frac{\nu_2}{k_{2x}^0}+\frac{\nu_3}{\varepsilon q_x^0}\right)\frac{\partial\tilde{E}_{2z}}{\partial t},\qquad(13)$$

where

$$E_{2z}=E_{1z}\,\frac{\exp(-q_x^0a)}{\sin(k_{2x}^0a)}\exp\left[-\frac{(\nu_1z-t)^2}{\tau^2}\right]\qquad(14)$$

[since the field (1), (2) only differs appreciably from zero in the interval $-a-q_x^{-1}<x<a+q_x^{-1}$, the inequalities used to derive formulas (13) imply that the pulse propagation time in this region is shorter than the pulse duration τ.]

It follows from Eqs. (13) and (14) that the signal envelope has a Gaussian profile and its group velocity is $v_{gr}=1/\nu_1$ (9). For a monochromatic wave ($\tau\to\infty$) expressions (13) and (14) are the same as formulas (1) and (2).

Let us now assume that an electron beam propagates in the field (1), i.e., on either side of the waveguide (note that this gives a weak dependence of the field amplitude E_{1z} on the z coordinate and time t). Fairly cumbersome calculations show that in this case the Maxwell equations can be reduced to the form

$$\frac{\partial P}{\partial z}+\nu_1\frac{\partial P}{\partial t}=-\int_a^{\infty}dx\int_{-l/2}^{l/2}dy\,\mathrm{Re}\cdot(\mathbf{j}\cdot\mathbf{E}^*).\qquad(15)$$

Here $\mathbf{j}$ is the electron beam current density and

$$P=\frac{c}{8\pi}\,\frac{l\omega^0k_z^0}{c\varepsilon(q_x^0)^3(k_{2x}^0)^2}\tilde{E}_{1z}^2\exp(-2q_x^0a)$$
$$\times\{\varepsilon[(q_x^0)^2+(k_{2x}^0)^2]+aq_x^0[(\varepsilon q_x^0)^2+(k_{2x}^0)]\}\qquad(16)$$

is the energy flux of the electromagnetic pulse across the xy plane. If the pulse duration is large ($\tau\to\infty$), expression (16) is the same as the energy flux of a monochromatic wave[6] (1), (2).

ELECTRON BEAM CURRENT IN THE FIELD OF A SURFACE WAVE

Let us assume that the initial electron beam has a Gaussian spread of momentum projections

$$f_0(\mathbf{p})=\left(\frac{4\ln2}{\pi}\right)^{3/2}\frac{1}{\Delta_\perp^2\Delta_\parallel}$$
$$\times\exp\left\{-4\ln2\,\frac{p_x^2+p_y^2}{\Delta_\perp^2}-4\ln2\,\frac{(p_z-p_0)^2}{\Delta_\parallel^2}\right\}.$$

We then direct the static magnetic field $\mathbf{H}(0,0-H_0)$ along this beam. We assume that L is the waveguide length and its beginning is located at the point $z=0$. Solving the Vlasov equation

$$\frac{\partial f}{\partial t}+\mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{p}}+e\left\{\mathbf{E}+\frac{1}{c}[\mathbf{v}\times(\mathbf{H}+\mathbf{H}_0)]\right\}\cdot\frac{\partial f}{\partial \mathbf{p}}=0$$

exactly for the static magnetic field and in the linear approximation for the field (1), we find the electron distribution function in these fields $f=f_0+f_1$ (Ref. 1). We then calculate the electron beam current $\mathbf{j}=e\rho_0\int\mathbf{v}\cdot f_1 d\mathbf{p}$. If the magnetic field is fairly strong, $H_0\gg mc\omega^0\Delta/|e|p_0$, we obtain

$$j_x=0,$$

$$j_z=j_{1z}+j_{2z}=-ie^2\rho_0\int d\omega\int_{-\infty}^{+\infty}dp_z\int_0^{\infty}p_\perp dp_\perp \times\int_0^{2\pi}d\varphi\frac{v_zE_z(\omega)}{\omega-k_zv_z+i\eta'}\frac{\partial f_0}{\partial p_z}\exp(-q_xx) \times\left\{\exp[i(k_zz-\omega t)]-\exp\left[i\left(\frac{\omega}{v_z}z-\omega t\right)\right]\right\} \quad (18)$$

(an infinitely small imaginary component $i\eta'$ is introduced in order to correctly bypass the Cherenkov pole).

Note that the first term in the current (18) is usually taken into account when the amplification region in a free electron laser is large[9] ($L\to\infty$) while the second term is taken into account when the interaction region is bounded.[10] Hence, in the first case the signal is amplified by electrons lying on the Cherenkov cone ($\omega-k_zv_z=0$), whereas in the second case it is amplified by electrons outside this cone ($\omega-k_zv_z\neq0$). We shall initially integrate over momentum and frequency in the current j_{1z}. We shall assume that the momentum $p_z=b$ satisfies the phase-matching condition $\omega-k_zv_z=0$ for $\omega=\omega^0$, $k_z=k_z^0$, and $p_\perp=0$. Then, the solution of this equation can be written in general form as follows $p_z=b+q_1p_\perp+q_2p_\perp^2+q_3\delta\omega$, where

$$q_1=0,\qquad q_2=\frac{b}{2m^2c^2},$$

$$q_3=\frac{b}{\omega^0\beta_0^2}\left(\frac{b}{mc}\right)^2(v_0\nu_1-1),\quad \beta_0=\frac{v_0}{c}. \quad (19)$$

Assuming for simplicity that $q_2\Delta_\perp^2\ll\Delta_\parallel$, we obtain

$$j_{1z}=\pi^{-1/2}(4\ln2)^{3/2}\rho_0r_0\lambda_0\left(\frac{p_0}{mc}\right)^2\frac{mcE_0(b-p_0)}{D_{fe}^3} \times\exp\left[-4\ln2\frac{(b-p_0)^2}{D_{fe}^2}\right]E_{1z} \times\exp\left[i(k_z^0-\omega^0t)-q_x^0x-\frac{\eta^2}{\tau^2}\frac{\Delta_\parallel^2}{D_{fe}^2}\right] \times[\cos(g_1\eta)+g_2\eta\sin(g_1\eta)]. \quad (20)$$

Here

$$D_{fe}=\left[\Delta_\parallel^2+16\ln(2)\frac{q_3^2}{\tau^2}\right]^{1/2} \quad (21)$$

is the effective width characterizing both the photon beam and the electron beam, $\eta=\nu_1z-t$,

$$g_1=16\ln(2)q_3(b-p_0)/\tau^2D_{fe}^2,\quad g_2=2q_3/\tau^2(b-p_0).$$

Note that the envelope of the current (20) is larger than the envelope of the amplified signal (13), (14) $\tau_j=\tau D_{fe}/\Delta_\parallel>\tau$. In addition, oscillations having the period $T_j=(\pi/8\ln2)\tau^2D_{fe}^2/q_3|b-p_0|$ appeared along the current envelope. We shall now calculate the current j_{2z}. We know[10] that in the short waveguide limit, we can neglect the electron momentum spread, i.e., we can assume that $f_0=\delta(p_x)\delta(p_y)\delta(p_z-p_0)$. Taking into account the expansion (8), we write the phase-matching condition in the form $\omega-k_zv_0=\omega^0-k_z^0v_0+\delta\omega(1-\nu_1v_0)$. Quite clearly, when the wave group velocity is equal to the electron beam velocity ($\nu_1^{-1}=v_0$) or the pulse duration is fairly large $\tau\gg|1-\nu_1v_0|/|\omega^0-k_z^0v_0|$, the second term in this expansion can be neglected. In this case, the current

$$j_{2z}=ie^2\rho_0E_{1z}\int d\mathbf{p}\frac{v_z}{\omega^0-k_z^0v_z}\frac{\partial f_0}{\partial p_z} \times\exp\left[i\omega^0(\nu_1z-t)-q_x^0x-\frac{(\nu_1z-t)^2}{\tau^2}\right] \quad (22)$$

is directly proportional to the height of the pulse being amplified (10), (13).

GAINS

Substituting the currents (20) and (22) into Eq. (15) and integrating the partial differential equation, we obtain $p=p_0\exp(G_{1,2}L)$, where

$$G_1=-4\sqrt{\pi}(4\ln2)^{3/2}\rho_0r_0\lambda_0Q\frac{mc}{p_0}\left(\frac{p_0}{D_{fe}}\right)^2\frac{b-p_0}{D_{fe}} \times\exp\left[-4\ln2\frac{(b-p_0)^2}{D_{fe}^2}\right]\exp\left(-\frac{\eta^2}{\tau^2}\frac{D_{fe}^2-\Delta_\parallel^2}{D_{fe}^2}\right) \times[\cos(g_1\eta)+g_2\eta\sin(g_1\eta)], \quad (23)$$

$$G_2=2\pi^2\rho_0r_0L^2\lambda_0^{-1}\beta_0^{-5}Q\left(\frac{mc^2}{r_0}\right)\frac{d}{d\Theta}\frac{\sin^2\Theta}{\Theta^2}. \quad (24)$$

Here the factor is

$$Q=(\varepsilon\beta_0^2-1)/(\varepsilon-1)\beta_0^2 \times\left\{1-2\pi\frac{a}{\lambda_0}\frac{1}{\varepsilon\beta_0^3}\frac{mc^2}{r_0}\left[1-\varepsilon\left(\frac{mc^2}{r_0}\right)^2\right]\right\},$$

and the parameter is $\Theta=(\omega^0-k_z^0v_0)L/2v_0$. In the first case, different points on the pulse envelope clearly have different gains $G_1=G_1(\eta)$, and this dependence is oscillatory. The gain depends on the electron energy spread and on the width of the signal frequency spread (21). For a monochromatic wave both these effects disappear.

We shall make a more detailed study of the case when the mismatch is $b-p_0=-D_{fe}/\sqrt{8\ln2}$. In this case, the center of the pulse envelope ($\eta=0$) has the maximum gain

$$G_{1\max}=8.4\rho_0 r_0\lambda_0 Q\,\frac{mc}{p_0}\left(\frac{p_0}{\Delta_\parallel}\right)^2. \tag{25}$$

We introduce the concept of the characteristic time τ_0 on the basis of the condition $\Delta_\parallel=4\sqrt{8\ln 2}\,q_3/\tau_0$ (20). Taking Eq. (19) into account, we obtain

$$\tau_0=2\pi^{-1}\sqrt{\ln 2}\,T_0|v_0\nu_1-1|\left(\frac{r_0}{mc^2}\right)^2\frac{p_0}{\Delta_\parallel}, \tag{26}$$

where $T_0=\lambda/c$.

We assume that the pulse duration is large if $\tau\gg\tau_0$. In this case, the effective width is $D_{fe}\approx\Delta_\parallel$ and the gain (25) is the same as the monochromatic wave gain[1]

$$G_0=8.4\rho_0 r_0\lambda_0 Q\,\frac{mc}{p_0}\left(\frac{p_0}{\Delta_\parallel}\right)^2. \tag{27}$$

If the pulse duration is short $T_0<\tau<\tau_0$, the gain (25) depends on the duration τ and decreases as it becomes shorter. For very short pulses ($\tau\ll\tau_0$) the gain of the envelope center is

$$G=\frac{\tau^2}{\tau_0^2}G_0. \tag{28}$$

Taking into consideration the definition (26), we find that the amplificati on of short pulses is determined only by their spectral width $\Delta\omega=1/\tau$.

We shall now analyze the gain at the pulse wings. Substituting Eq. (27) into Eq. (23), we obtain

$$G_1(\eta)=G_0\exp\left(-\frac{\eta^2}{\tau^2}\,\frac{D_{fe}^2-\Delta_\parallel^2}{D_{fe}^2}\right)[\cos(g\eta)+g\eta\sin(g\eta)]. \tag{29}$$

Here we have

$$g=-\frac{4\sqrt{2\ln 2}\,q_3}{\tau^2 D_{fe}}=-\frac{\sqrt{2}\,\tau_0}{\tau^2(1+\tau_0^2/\tau^2)^{1/2}}. \tag{30}$$

Clearly the trigonometric factor in expression (29) decreases from one to zero as $|\eta|$ increases. The gain then becomes negative, i.e., the electron beam absorbs the energy of these sections of the wave. Then G_1 again goes to zero and so on. Equating the expressions in parentheses to zero, we obtain the condition for which the gain goes to zero: $\tan g\eta=-1/g\eta$. Quite clearly, by using the oscillatory nature of the gain (29) (especially in regimes with the mismatch $b-p_0>0$) we can influence the shape of the pulse envelope.

If we need to amplify an electromagnetic pulse without changing its shape, the second regime (24) should be used.

CONCLUSIONS

We shall illustrate these results by means of numerical estimates, first considering the case of a long waveguide. We shall assume that the electron beam density is $\rho_0=0.5\times10^9\ \mathrm{cm}^{-3}$ ($j_0=1.25\ \mathrm{A/cm}^2$), its average energy is $E_0=U+mc^2=660\ \mathrm{keV}$, and the angular and energy spreads are $\delta=\Delta_\perp/p_0=10^{-2}$, and $\Delta/E_0=0.5\times10^{-2}$. We shall assume that a pulse having the carrier frequency $\omega^0=4.7\times10^{-11}\ \mathrm{Hz}$ ($\lambda_0=4\ \mathrm{mm}$) is amplified. Suppose that the waveguide is made of quartz ($\varepsilon=3.8$), its length is $L_1=7\ \mathrm{cm}$, and the static magnetic field is $H_0=4\ \mathrm{G}$. We then find from Eq. (26) that for the selected parameters the characteristic time is $\tau_0=0.5\ \mathrm{ns}$. If the pulse duration is large, $\tau=3$, $\tau_0=1.5\ \mathrm{ns}$, the gain (27) is $G_0=0.1\ \mathrm{cm}^{-1}$. The gain (28) for a short pulse $\tau=\tau_0/3=0.2\ \mathrm{ns}$ is an order of magnitude lower $G_1=0.01\ \mathrm{cm}^{-1}$. This factor should be taken into account when setting up an experiment. Note that the gain (29) goes to zero for the first and second times at the points $|\eta_1|=0.2\ \mathrm{ns}$, $|\eta_2|=0.5\ \mathrm{ns}$.

We shall now analyze the second regime (24). We shall assume that the waveguide length $L_2=4\ \mathrm{cm}$ and that all the other waveguide and radiation parameters are the same as in the previous case. Bearing in mind that the function $d\sin^2\Theta/d\Theta^2$ reaches its maximum of 0.5 for $\Theta=-1.26$, we find that in this case the average electron beam energy should be $E_0=690\ \mathrm{keV}$. Note that under these conditions the pulse group velocity $v_{\mathrm{gr}}/c\approx1/\varepsilon\beta_0=0.39$ is not equal to the electron velocity $v_0/c=0.67$. Assuming that the electron density is $\rho_0=0.6\times10^{10}\ \mathrm{cm}^{-1}$ ($j_0=15\ \mathrm{A/cm}^2$), we then obtain the gain $G_2=0.1\ \mathrm{cm}^{-1}$ for pulses of duration $\tau>56\ \mathrm{ps}$.

This work was supported by the International Scientific and Technical Center, Grant No. A-87.

[1] V. M. Harutunian and S. G. Oganesyan, Phys. Rep. **270**, 217 (1996).
[2] J. E. Walsh, Adv. Electron. Electron Phys. **58**, 271 (1982).
[3] J. E. Walsh and J. B. Murphy, IEEE J. Quantum Electron. **QE-18**, 1259 (1982).
[4] E. P. Garate, J. E. Walsh, C. H Shaughnessy *et al.*, Nucl. Instrum. Methods Phys. Res. A **259**, 125 (1987).
[5] E. P. Garate, C. H. Shaughnessy, and J. E. Walsh, IEEE J. Quantum Electron. **QE-23**, 1627 (1987).
[6] R. A. Akopov, E. M. Laziev, and S. G. Oganesyan, Zh. Tekh. Fiz. **65**(1), 99 (1995) [Tech. Phys. **40**, 52 (1995)].
[7] N. K. Zhevago and V. N. Glebov, Zh. Éksp. Teor. Fiz. **111**, 847 (1997) [JETP **84**, 466 (1997)].
[8] S. G. Oganesyan, Pis'ma Zh. Tekh. Fiz. **22**(7), 32 (1996) [Tech. Phys. Lett. **22**(4), 279 (1996)].
[9] R. H. Pantell, G. Soncini, and H. E. Putthoff, IEEE J. Quantum Electron. **QE-4**, 905 (1968).
[10] V. P Sukhatme and P. A. Wolf, IEEE J. Quantum Electron. **QE-10**, 870 (1974).

Translated by R. M. Durham

Natural electromagnetic modes of a resonator formed by cutoff waveguides

Yu. G. Makeev and A. P. Motornenko

Institute of Radio Engineering and Electronics, Academy of Sciences of Ukraine, 310085 Kharkov, Ukraine
(Submitted June 24, 1997)
Zh. Tekh. Fiz. **69**, 84–88 (April 1999)

A spectral three-dimensional problem is solved to determine the resonant frequencies of a branch formed by the orthogonal intersection of cutoff rectangular waveguides. Expressions for the H_{mng} and E_{mng} modes are derived and analyzed. Two classes of free nonradiating modes typical of this structure are identified and analyzed. Numerical calculations are made of the dispersion relations obtained for the different modes. © *1999 American Institute of Physics.* [S1063-7842(99)01504-4]

INTRODUCTION

The electrodynamic characteristics of waveguide branches, variable cross-section waveguides, T-junctions, abruptly expanding sections, and so on have been studied by various authors.[1–6] Resonators based on waveguide branches have the property, which is of practical importance, that they can support natural modes of oscillation when the waveguides are cutoff waveguides. This feature is characteristic of waveguide structures formed from rectangular[7,8] and circular waveguides.[9,10] In these studies, which were concerned with damped modes, free oscillations were studied for the two-dimensional case.

The sparse mode spectrum, high intrinsic Q-factors (~7000 or higher), and the possibility of localizing the electromagnetic field in small volumes mean that these resonators can have extensive practical applications. At present, waveguide branches are used as measuring sections to determine the local parameters of dielectric and ferrite materials,[10] to determine the direction of the axes of the refractive index ellipsoid for anisotropic materials, and so on. These branches are also used to fabricate passive and active microwave devices. Astionenko and Motornenko[11] described a semiconductor microwave oscillator with a wide mechanical frequency tuning band and the capacity to sum the power of several semiconductor diodes.

The fields of application of these resonant structures can be extended by studying the characteristics of the three-dimensional modes in them. Makeev *et al.*[12,13] studied free three-dimensional oscillations in an orthogonal branch of circular and radial waveguides. Here we describe a method of calculating the resonant frequencies of free electromagnetic oscillations under conditions where the field varies along three axes of the coordinate system in an orthogonal branch of rectangular waveguides.

FORMULATION AND SOLUTION OF THE SPECTRAL PROBLEM

We shall solve the problem of free oscillations in the structure being investigated (Fig. 1) assuming that the damped $E_{m(n)g}$ and $H_{m(n)g}$ modes exist. We shall use the method of partial regions for which we isolate three regions in the waveguide branch: *1* and *2* are regions of rectangular waveguides of cross sections $b\times c$ and $a\times c$, respectively, and *3* is a region of dimensions $a\times b\times c$ formed by the intersection of waveguides *1* and *2*. We shall assume that each region is filled with a magnetic insulator of permittivity ε_i and magnetic permeability μ_i, where $i=1$, 2, or 3. To simplify the problem, we neglect the losses of electromagnetic energy in the metal walls and the filling material. The origin is located at the center of the branch, as shown in Fig. 1. We assume that waveguides *1* and *2* are unbounded in the direction of the *x* and *y* axes. Using the symmetry property of the branch relative to the center of the coordinate system, we select even $(m,n=0,2,\ldots)$ and odd $(m,n=1,3,\ldots)$ resonant modes relative to the *x* and *y* axes, where *m* and *n* are the number of variations of the field along the corresponding walls of the waveguides forming the branch. Here we consider modes with an odd number of variations of the field along the *x* and *y* axes. The number of variations of the field (g) along the *z* axis is arbitrary.

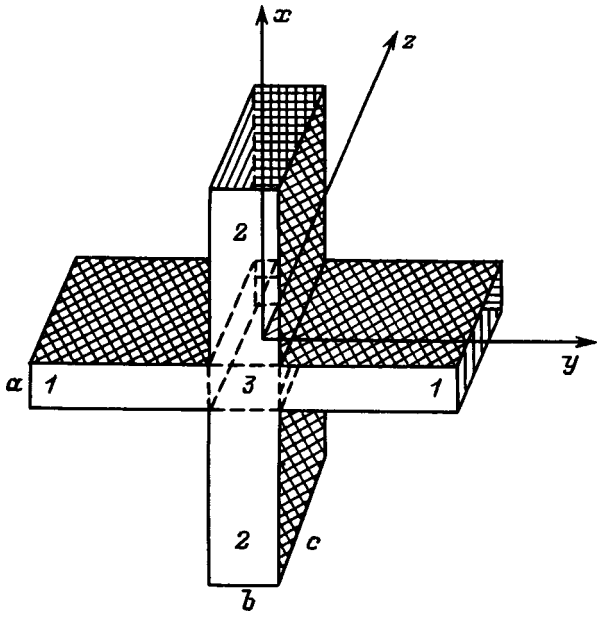

FIG. 1. Waveguide branch formed by two orthogonal waveguides.

In order to describe the electromagnetic field in all regions of the branch, we shall use the magnetic ($\mathbf{\Pi}^h$) and electric ($\mathbf{\Pi}^e$) Hertz vectors for *H*-type and *E*-type waves, respectively. The components of the electromagnetic field required to solve the problem are expressed by means of Maxwell equations in terms of the Hertz vectors using a well-known method.[1]

The magnetic and electric Hertz vectors for region *1* are written as Fourier series in terms of the damped H_{mg} or E_{mg} electromagnetic modes of the corresponding waveguide

$$\mathbf{\Pi}^{h_1}_{mg}=-\mathbf{z}_0\sum_{m,g}A_{mg}\frac{1}{p_m^2p_g^2}\sin p_{mx}\times\exp[-\gamma_{mg}(y-b/2)]\begin{cases}\sin p_gz,\\ \cos p_gz,\end{cases}\tag{1}$$

where

$$\gamma_{mg}=(p_m^2+p_g^2-k^2\varepsilon_1\mu_1)^{1/2}\begin{cases}g=0,\ 2,\ 4,\ldots,\\ g=1,\ 3,\ 5,\ldots,\end{cases}$$

$$\mathbf{\Pi}^{e_1}_{mg}=-\mathbf{z}_0\frac{\omega\varepsilon_1}{i}\sum_{m,g}C_{mg}\frac{1}{p_m^2p_g^2}\cos p_{mx}\times\exp[-\gamma_{mg}(y-b/2)]\begin{cases}\sin p_gz,\\ \cos p_gz,\end{cases}\tag{2}$$

where

$$\gamma'_{mg}=(p_m^2+p_g^2-k^2\varepsilon_1\mu_1)^{1/2}\begin{cases}g=1,\ 3,\ 5,\ldots,\\ g=0,\ 2,\ 4,\ldots,\end{cases}$$

$A_{mg}(C_{mg})$ are the amplitudes of the magnetic (electric) modes in the first rectangular waveguide, $p_m=m\pi/a$, $p_g=g\pi/c$, $\mathrm{k}=2\pi/\lambda$, λ is the resonance wavelength, ω is the resonant angular frequency, $i=\sqrt{-1}$, and $\mathbf{z}_0$ is the unit vector along the *z* axis.

By analogy with the previous case, the Hertz vectors of the second waveguide (region *2*) are written as the sum of the damped waves in the *x* direction

$$\mathbf{\Pi}^{h_2}_{ng}=-\mathbf{z}_0\sum_{n,g}B_{ng}\frac{1}{p_n^2p_g^2}\sin p_ny\times\exp[-\gamma_{ng}(x-a/2)]\begin{cases}\sin p_gz,\\ \cos p_gz,\end{cases}\tag{3}$$

$$\mathbf{\Pi}^{e_2}_{ng}=-\mathbf{z}_0\frac{\omega\varepsilon_2}{i}\sum_{n,g}D_{ng}\frac{1}{p_n^2p_g^2}\cos p_ny\times\exp[-\gamma_{ng}(x-a/2)]\begin{cases}\sin p_gz,\\ \cos p_gz,\end{cases}\tag{4}$$

where $B_{ng}(D_{ng})$ are the amplitudes of the magnetic (electric) modes in the second waveguide, and $p_n=n\pi/b$.

The expressions for γ_{ng} and γ'_{ng} are the same as those given for γ_{mg} and γ'_{mg} in the first region with *m* replaced by *n* and $\varepsilon_1\mu_1$ replaced by $\varepsilon_2\mu_2$.

The Hertz vectors describing the electromagnetic field in the third region are expressed as the superposition of the Hertz vectors of the first and second regions[6]

$$\mathbf{\Pi}^{h_3}_{m(n)g}=-\mathbf{z}_0\sum_{m,g}A'_{mg}\frac{1}{p_m^2p_g^2}\cos p_mx\frac{\sinh(\Gamma_{mg}y)}{\cosh(\Gamma_{mg}b/2)}\times\begin{cases}\sin p_gz\\ \cos p_gz+\mathbf{z}_0\sum_{n,g}B'_{ng}\frac{1}{p_n^2p_g^2}\sin p_ny\end{cases}\times\frac{\sinh(\Gamma_{ng}x)}{\cosh(\Gamma_{ng}a/2)}\begin{cases}\sin p_gz,\\ \cos p_gz,\end{cases}\tag{5}$$

$$\mathbf{\Pi}^{e_3}_{m(n)g}=-\mathbf{z}_0\frac{\omega\varepsilon_3}{i}\sum_{m,g}C'_{mg}\frac{1}{p_m^2p_g^2}\cos p_mx\frac{\cosh(\Gamma_{mg}y)}{\cosh(\Gamma_{mg}b/2)}\times\begin{cases}\sin p_gz\\ \cos p_gz-\mathbf{z}_0\frac{\omega\varepsilon_3}{i}\sum_{n,g}D'_{ng}\frac{1}{p_n^2p_g^2}\sin p_ny\end{cases}\times\frac{\cosh(\Gamma_{ng}x)}{\cosh(\Gamma_{ng}a/2)}\begin{cases}\sin p_gz,\\ \cos p_gz,\end{cases}\tag{6}$$

where

$$\Gamma_{mg}=(p_m^2+p_g^2-k^2\varepsilon_3\mu_3)^{1/2}\begin{cases}g=0,\ 2,\ 4,\ldots,\\ g=1,\ 3,\ 5,\ldots,\end{cases}$$

$$\Gamma'_{ng}=(p_n^2+p_g^2-k^2\varepsilon_3\mu_3)^{1/2}\begin{cases}g=1,\ 3,\ 5,\ldots,\\ g=0,\ 2,\ 4,\ldots\ .\end{cases}$$

The expressions for the Hertz vectors in each region are written in a form satisfying the Helmholtz wave equation. The electromagnetic field expressed in these regions in terms of the appropriate values of the magnetic and electric Hertz vectors satisfies the radiation condition and the finite energy condition in any bounded region of the structure.[4]

As a result of imposing the conditions of continuity at the interfaces on the tangential components of the magnetic and electric fields, we obtain a system of functional equations which can be reduced to two systems of linear algebraic equations of the second kind for the coefficients $A_{kg}, A_{mg}(C_{kg}, C_{mg})$ (where *k* is the number of variations of the electromagnetic field of a resonant mode in an empty waveguide) by projecting the magnetic and electric modes of first one and then the other waveguide in terms of eigenfunctions. The difference between these systems is that in one *g* is even and in the other it is odd. Thus, these systems can be combined by assuming $g=0,\ 1,\ 2,\ 3,\ldots$, and then written as follows:

$$C_{kg}-\frac{16\varepsilon_3^2}{ab}\sum_m C_{mg}\frac{(-1)^{(m-1)/2}p_k^2}{S'_mf'_{kg}}Q'_{mg}=0,\tag{7}$$

$$A_{kg}-\frac{16}{\mu_3^2 ab}\sum_m A_{mg}\frac{(-1)^{(m-1)/2}p_k^2}{S_m f_{kg}}Q_{mg}=0, \qquad (8)$$

where

$$S_{m(n)}=\Gamma^2_{m(n)g}+p^2_{n(m)}; \quad S'_{m(n)}=\Gamma'^2_{m(n)g}+p^2_{n(m)};$$

$$f_{kg}=(\mu_1\alpha_{31}\gamma_{kg})^{-1}+(\mu_3\Gamma_{kg})^{-1}\tanh(\Gamma_{kg}b/2);$$

$$f_{ng}=(\mu_2\alpha_{32}\gamma_{ng})^{-1}+(\mu_3\Gamma_{ng})^{-1}\tanh(\Gamma_{ng}a/2);$$

$$f'_{kg}=\varepsilon_1\alpha_{31}\gamma'_{kg}+\varepsilon_3\Gamma'_{kg}\tanh(\Gamma_{kg}b/2);$$

$$f'_{ng}=\varepsilon_2\alpha_{32}\gamma'_{ng}+\varepsilon_3\Gamma'_{ng}\tanh(\Gamma_{ng}a/2);$$

$$Q_{mg}=\sum_n \frac{(-1)^{(n-1)/2}}{S_n f_{ng}}; \quad Q'_{mg}=\sum_n \frac{p_n^2(-1)^{(n-1)/2}}{S'_n f'_{ng}};$$

$$\alpha_{31}=\frac{k^2\varepsilon_3\mu_3+p_kp_g-p_g^2}{k^2\varepsilon_1\mu_1+p_kp_g-p_g^2}; \quad \alpha_{32}=\frac{k^2\varepsilon_3\mu_3+p_np_g-p_g^2}{k^2\varepsilon_1\mu_1+p_np_g-p_g^2}.$$

The solution of the systems (7) and (8) is nontrivial if the determinant of each system is zero. By setting the determinants of the systems of linear algebraic equations (7) and (8) to zero, we obtain dispersion relations which can be used to find the spectral parameters of the structure using its given geometric dimensions and the known parameters of magnetic insulators in each region of the branch.

ANALYSIS OF RESULTS

We shall begin our analysis of the system of equations (7) and (8) with the case where the free oscillations of the branch have no variations of the field along the z axis (two-dimensional modes). Assuming that $g=0$ in Eqs. (7) and (8), we obtain systems of equations which are the same as those described by Korobkin *et al.*[7,8] It should be noted that the structure of the system (7) is identical to the system studied in detail by Shestopalov *et al.*[4] Thus, the conclusions reached in that study that the method of reduction can be applied to the determinants of the equations can also be taken to apply to the equations derived here. The modes under consideration in the waveguide branch are described by electromagnetic fields which decay exponentially with increasing distance from the branch region. Thus, Eqs. (7) and (8) contain no matrix quantities (such as $S_{m(n)}$ and $S'_{m(n)}$) which lead to singularity of the solutions.

The free oscillations described by Eq. (7) will subsequently be denoted as H_{mng} modes and those described by Eq. (8) will be denoted as E_{mng} modes.

H_{mng} MODES IN AN ORTHOGONAL BRANCH

For the H_{mng} modes, as in Refs. 7 and 8, we can isolate two classes of resonant nonradiating modess: natural modes in the branch, with real propagation constants γ_{mg}, γ_{ng}, Γ_{mg}, and Γ_{ng} in all regions, and natural waveguide–insulator modes for which γ_{mg} and γ_{ng} are real and Γ_{mg} and Γ_{ng} are imaginary.

A distinguishing feature of the first class of modes is that they can exist in an empty branch, whereas the second class can only occur when the branch contains an insulator or a magnetic insulator which provides conditions for wave propagation in this region.

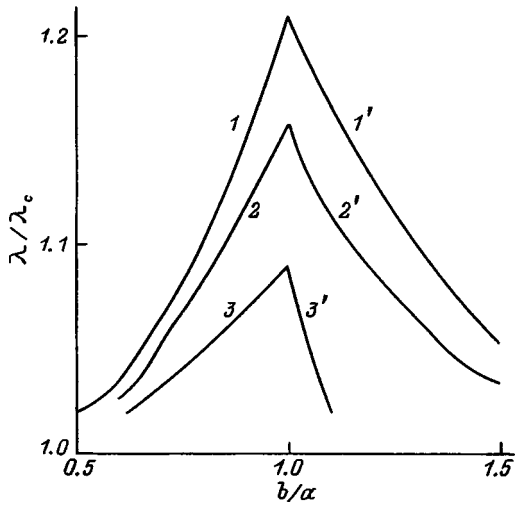

FIG. 2. Calculated curves of the normalized resonant wavelength for the H_{110} (*1, 1′*) and H_{111} (*1, 1′–3, 3′*) modes as a function of b/a: $\varepsilon_1=\varepsilon_2=\varepsilon_3=1$, $\mu_1=\mu_2=\mu_3=1$; $a/c=0$ (*1, 1′*), 0.5 (*2, 2′*), and 1.0 (*3, 3′*).

Figure 2 gives the calculated curves of the normalized resonant wavelength of the first class of H_{110} and H_{111} modes in an empty branch plotted as a function of b/a for various values of a/c. The calculations were made assuming that one damped wave exists in each waveguide forming the branch. By analogy with Ref. 7, the values of the resonant wavelength are normalized to the critical wavelength in the waveguide section having the largest transverse dimensions, i.e., curves *1–3* are normalized to $\lambda_c=2/[(1/a)^2+(1/c)^2]^{1/2}$, and curves *1′–3′* are normalized to $\lambda_c=2/[(1/b)^2+(1/c)^2]^{1/2}$. The points of intersection of curves *1, 1′* (*2, 2′* and *3, 3′*) correspond to the case where the propagation constants are the same in both waveguides. Calculations of Eq. (7) for the H_{110} mode and the similar equation for the H_{111} mode in the approximation indicated above showed that for $a/c=0$ ($c\to\infty$) the resonant frequencies of these modes are the same for all b/a and they are also the

TABLE I.

	λ/λ_c	
	Mode	
Number of terms in inner sum	H_{110}	H_{111}
1	1.211617	1.090702
2	1.206781	1.088932
3	1.207831	1.089318
4	1.207445	1.089176
5	1.207672	1.089243
6	1.207527	1.089206
7	1.207588	1.089229
8	1.207548	1.089214
9	1.207575	1.089224
10	1.207556	1.089217

TABLE II.

Order of determinant of sum	λ/λ_c Mode H_{110}	λ/λ_c Mode H_{111}
1	1.211617	1.090702
2	1.204085	1.087940
3	1.205039	1.088359
4	1.204612	1.088134
5	1.204765	1.088191
6	1.204657	1.088151
7	1.204708	1.088170
8	1.204666	1.088155
9	1.204688	1.088163
10	1.204667	1.088155

same for all H_{11g} modes. Thus when the dimension c is increased appreciably, the spectrum of natural mode frequencies of the branch becomes more abundant. An increase in the ratio a/c (curves *2, 2′* and *3, 3′*) leads to a decrease in the resonant wavelength of the H_{111} mode, while that of the H_{110} mode remains constant.

In order to determine whether the infinite system of equations converges, we calculated the determinant of the system (7) as a function of the order of the determinant and the number of terms in the inner sum in Eq. (7) as a function of the order of the determinant and the number of terms in the inner sum for the H_{110} and H_{111} modes. The parameters of the system used for the calculations were: $a/c=1$, $a/b=1$, $\varepsilon_1=\varepsilon_2=\varepsilon_3=1$, $\mu_1=\mu_2=\mu_3=1$. Table I gives the results which show that λ/λ_c varies as a function of the number of terms N taken into account in the inner sum in Eq. (7). In this case, the order of the determinant (M) was 1. Dependences of λ/λ_c on the order of the determinant are illustrated in Table II. For the calculations the number of terms N in the inner sum and the order of the determinant M were taken to be the same.

Figure 3 gives the results of numerical calculations of the dependence $\lambda/\lambda_c=f(\varepsilon)$ $(\lambda_c=2\sqrt{\varepsilon}/[(1/a)^2+(1/c)^2]^{1/2})$ which illustrate how an insulator of dimensions $a\times b\times c$ in the region of intersection of the waveguides influences the resonance properties of the branch. The conditions used to calculate the curves of λ/λ_c are the same as in the previous case. For clarity we consider the case $a=b=c$. Curves *1* and *2* in the range $\lambda/\lambda_c\geqslant 1$ correspond to the first class of H_{110} and H_{111} natural modes. Part of curve *2* (for $\lambda/\lambda_c<1$) corresponds to the H_{111} waveguide–insulator mode. It can be seen from Fig. 3 that the the longest wavelengths are for the natural modes in the branch when all the regions are cutoff regions.

E_{mng} MODES IN AN ORTHOGONAL BRANCH

As above, we shall separate the natural modes of the branch with real propagation constants and the natural waveguide–insulator modes. An analysis of Eq. (8) for the E_{mng} modes showed that no natural modes of the first class

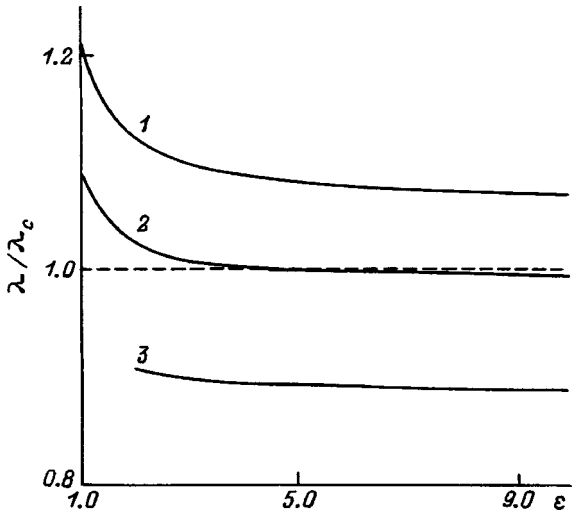

FIG. 3. Calculated curves of the normalized resonant wavelength as a function of the permittivity of the insulator filling the branch region: *1(2)*—$H_{110}(H_{111})$ mode of branch, *3* — E_{111} mode: λ_c is the critical wavelength in the waveguide with the insulator.

exist in the orthogonal branch, just as in the two-dimensional case,[8] where a similar result was obtained for the E_{mn0} mode. Thus, we made numerical calculations for the fundamental E_{111} waveguide–insulator mode. In the calculations we assumed that there is a single type of propagating mode in the waveguide intersection region and we analyzed the case $a=b=c$. Curve *3* in Fig. 3 gives the normalized resonant wavelength λ/λ_c as a function of the permittivity of the sample filling the waveguide intersection region for the E_{111} mode.

An orthogonal branch of rectangular waveguides was investigated experimentally for the case $a=b=c$ in the 2–3 cm wavelength range. The resonant frequencies of the H_{110} and H_{111} modes were measured in the empty branch of rectangular waveguides of 11×11 mm cross section. The difference between the calculated and measured values of the resonant frequencies did not exceed 1%. For the practical implementation of devices based on these structures it is best to use the lowest-frequency H_{110} or H_{111} modes.

CONCLUSIONS

A spectral problem of practical importance has been solved for an electrodynamic system in the form of an orthogonal branch of rectangular cutoff waveguides. The results can be used to develop active and passive components for microwave engineering and also to develop means of making nondestructive measurements of the parameters of ferrite and dielectric materials or for monitoring these in a production cycle. The method used here can also be applied to study the resonant characteristics of structures in other fields of physics (such as acoustics and gasdynamics).

[1] R. Mittra and S. W. Lee, *Analytical Techniques in the Theory of Guided Waves* [Macmillan, New York (1971); Mir, Moscow (1974), 324 pp.].
[2] N. N. Voĭtovich, B. Z. Katsenelenbaum, and A. N. Sivov, *Generalized Method of Natural Modes in Diffraction Theory* [in Russian], Nauka, Moscow (1977), 416 pp.
[3] A. S. Il'inskiĭ and G. Ya. Slepyan, *Oscillations and Waves in Electrody-*

namic Systems with Losses [in Russian], Moscow State University Press, Moscow (1983), 232 pp.
[4] V. P. Shestopalov, A. A. Kirilenko, and L. A. Rud', *Resonant Scattering of Waves*, Vol. 2 [in Russian], Naukova Dumka, Kiev (1986), 215 pp.
[5] *Computer Techniques for Electromagnetics*, edited by R. Mittra [Pergamon Press, Oxford (1973); Mir, Moscow (1977), 485 pp.].
[6] E. Kuhn, AEU No. **27,** 497 (1973).
[7] V. A. Korobkin and V. V. Osintsev, Zh. Tekh. Fiz. **55**, 1907 (1985) [Sov. Phys. Tech. Phys. **30**, 1121 (1985)].
[8] V. A. Korobkin, E. V. Obol'yaninova, and V. V. Osintsev, Radiotekh. Elektron. **34**, 1770 (1989).
[9] V. A. Korobkin and Yu. G. Makeev, Radiotekh. Elektron. **32**, 526 (1987).
[10] V. A. Korobkin and Yu. G. Makeev, Inventor's Certificate No. 1117538 (USSR), International Classification G01R 27/26 G01N 22/02; pub. Byull. Izobret. No. 37 (1984) (Priority date Feb. 15, 1980).
[11] I. A. Astionenko and A. P. Motornenko, Radiotekh. Elektron. **7**, 1398 (1991).
[12] Yu. G. Makeev, L. A. Rud', and S. Yu. Ostritskaya, Radiotekh. Elektron. **39**, 1497 (1994).
[13] Yu. G. Makeev and A. P. Motornenko, Zh. Tekh. Fiz. **64**(9), 118 (1994) [Tech. Phys. **39**, 924 (1994)].

Translated by R. M. Durham

Modeling and design of a magnetic resonance mass spectrometer with a resolution of $\sim 10^6$

N. N. Aruev and E. L. Baĭdakov

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted September 5, 1997)
Zh. Tekh. Fiz. **69**, 89–96 (April 1999)

The equations of motion for a charged particle in crossed magnetic and electric fields are used to make a numerical simulation of the operation of a magnetic resonance mass spectrometer for which the ion trajectory in the analyzer is divided into various sections. The calculations take into account the real structural dimensions and the parameters of the main components and systems and also allow for its operating characteristics in various regimes, as obtained by investigating a prototype. The calculation s yield the optimum geometric, electrical, frequency, phase, and magnetic parameters of the device, for which one can achieve a resolution $R_{5096} \approx 1.35 \times 10^6$ and a current efficiency $K_I \approx 0.006$. © *1999 American Institute of Physics.* [S1063-7842(99)01604-9]

1. The magnetic resonance mass spectrometers (MRMSs) proposed[1] and developed[2–4] at the A. F. Ioffe Physicotechnical Institute of the Russian Academy of Sciences possess good analytical characteristics which enable these devices to be used to solve various physical problems. One line of development involves maximizing the sensitivity at the expense of some loss of resolution. For instance, MRMSs having the highest absolute and isotopic sensitivities recorded so far have been used to study the isotopic composition of helium and other rare gases: primary helium, neon, and argon have been discovered in the Earth's mantle;[5] ideas have been developed on the degassing of the Earth;[6] the rate of sedimentation and age of sedimentary rocks have been determined, especially the growth rate of manganese iron nodules,[7] and the half-life of tritium has been measured by a new helium isotope method,[8] and so on.

Another line of development of MRMSs is aimed at achieving high resolution with correspondingly lower sensitivity. These devices are required for precision measurements, e.g., to measure atomic masses or fundamental physical constants. Magnetic resonance mass spectrometers have been used for the most accurate measurements of the proton magnetic moment in nuclear magnetons[9,10] μ_p'/μ_N and have also recorded the $^3H^+ - ^3He^+$ doublet for the first time, in which the difference between the ion mass is $\Delta M \approx 2 \times 10^{-5}$ u (Ref. 11). The resolution of this mass spectrometer[11] at the half-height of the mass peak was $R_{50\%} \approx 3.5 \times 10^5$. Since this MRMS was based on a permanent magnet with a magnetic induction $B \approx 0.12$ T, it operated in a narrow range of mass (3–4 u) and could not be used for a wide range of applications. Nevertheless, this mass spectrometer was used to verify experimentally and confirm the validity of the proposed MRMS theory, it was used to investigate various operating regimes, and to study how changes in the electrical, frequency, and magnetic parameters influence the analytical characteristics.[12,13]

Taking these results as our basis, we began to design and construct an MRMS with a rated resolution at half-height of the mass peak greater than 10^6, based on a specially developed precision electromagnet for which the magnetic induction can vary between 0.05 and 0.5 T with a measurable mass range of 3–200 u. As well as having high resolution, this spectrometer should have the highest possible stability and operate stably under various conditions. This device is designed for precision measurements of the atomic masses of stable and unstable isotopes over a wide range of mass, measurements of the mass difference of the $^3H^+ - ^3He^+$ doublet to determine the rest mass of an electron antineutrino, determining the energies of chemical bonds in molecules, and measurements of fundamental physical constants, in particular μ_p'/μ_N, and so on.

2. The operating principle of the MRMS, its various operating regimes, and proposed methods of calculating its analytical characteristics were described in detail in Refs. 10, 12, and 13. Thus, we shall merely give a schematic of the MRMS analyzer (Fig. 1) using the notation adopted in the earlier publications. As in these previous studies, the program for simulating the MRMS analyzer is based on calculating the trajectory parameters of an isolated ion in the central plane of the device. Whereas Mamyrin *et al.*[11] used nine transcendental equations to describe the ion motion in the analyzer and Aruev *et al.*[12] used of the order of fifty equations, we use what we consider to be a more logical method of simulating and calculating the parameters by solving the equations of motion for an ion in the device. For this purpose the ion trajectory in the analyzer is divided into various sections, where the ions are exposed to the action of the static magnetic field alone or to superposed static magnetic and alternating electric fields. In an analyzer with a three-chamber modulator there are ten such sections: $S_0S_{A1}S_1$ — between the source and the modulator, d_1 — in the first modulator gap, d_2 — in the central electrode of the modulator (an arbitrary field-free space), d_3 — in the second modulator gap, S_1S_2 — from the modulator to the drift slit,

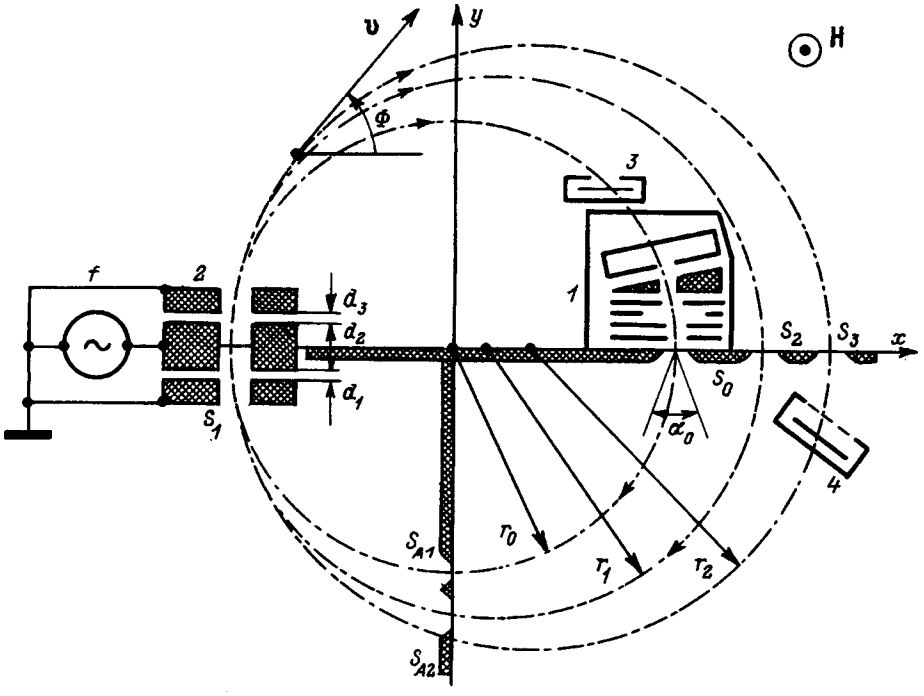

FIG. 1. Schematic of MRMS analyzer: *1* — ion source, *2* — modulator, *3* — ion collector, *4* — reflecting gap, *f* — rf modulating voltage generator, S_0 — source exit slit, S_1 — slits in modulator electrodes, S_2 — drift slit, S_3 — analyzer exit slit, S_{A1} and S_{A2} — aperture slits, d_1 and d_3 — widths of gaps between central and side electrodes in modulator, and d_2 — width of central modulator electrode.

$S_2S_{A2}S_1$ — from the drift slit to the modulator, d_1, d_2, d_3 — second transit of the modulator, and S_1S_3 — from the modulator to the analyzer exit slit. The results of the numerical calculations for each section of trajectory are the initial conditions for the next section.

We shall position the x and y axes of the coordinate system in the central plane of the device so that the magnetic field vector will be directed along the z axis (Fig. 1). We shall assume that the modulator gap is an ideal plane-parallel capacitor perpendicular to the central plane of the device. The system of equations describing the ion motion in the crossed uniform magnetic and alternating electric fields in the planar modulator gap is then given by

$$\frac{d^2x}{dt^2}=\omega_0\frac{dy}{dt}+D(\omega_0^2-\omega^2)\sin\Theta\,\sin\omega t,$$

$$\frac{d^2y}{dt^2}=-\omega_0\frac{dx}{dt}-D(\omega_0^2-\omega^2)\cos\Theta\,\sin\omega t \qquad (1)$$

with the initial conditions at time t_n

$$x(t_n)=x_n,\quad y(t_n)=y_n,\quad \dot{x}(t_n)=\dot{x}_n,\quad \dot{y}(t_n)=\dot{y}_n.$$

Here $x(t)$ and $y(t)$ are the coordinates of an ion of charge q and mass m at time t, $\omega_0=qH_0/mc$ is the ion cyclotron frequency in the uniform magnetic field H_0, $|E|=(U/d)\sin\omega t$ is the field strength in the modulator gap, U is the amplitude, and ω is the angular frequency of the modulating voltage;

$$D=\frac{qU}{md(\omega_0^2-\omega^2)},$$

d is the width of the modulator gap, and Θ is the angle formed by the plane of the modulator gap and the x axis.

The solution is given by

$$\begin{aligned}x(t)=&x_C+r\,\sin[\omega_0(t-t_n)-\phi]\\&+D\left\{\frac{\omega}{\omega_0}\cos\omega t_n(\cos\Theta-\cos[\omega_0(t-t_n)-\Theta])\right.\\&+\frac{\omega_0}{\omega}\cos\Theta(\cos\,\omega t-\cos\,\omega t_n)\\&\left.+\sin[\omega_0(t-t_n)-\Theta]\sin\,\omega t_n+\sin\Theta\,\sin\,\omega t\right\},\end{aligned}$$

$$\begin{aligned}y(t)=&y_C+r\,\cos[\omega_0(t-t_n)-\phi]\\&+D\left\{\frac{\omega}{\omega_0}\cos\omega t_n(\sin\Theta+\sin[\omega_0(t-t_n)-\Theta])\right.\\&+\frac{\omega_0}{\omega}\sin\Theta(\cos\,\omega t-\cos\,\omega t_n)\\&\left.+\cos[\omega_0(t-t_n)-\Theta]\sin\,\omega t_n+\cos\Theta\,\sin\,\omega t\right\},\end{aligned} \qquad (3)$$

where (x_C,y_C) are the coordinates of the center of the circle of radius r along which the ion would propagate in a uniform magnetic field in the absence of an electric field

$$x_C=x_n+\dot{y}_n/\omega_0,\quad y_C=y_n-\dot{x}_n/\omega_0, \qquad (4)$$

$|v_n|=\sqrt{\dot{x}_n^2+\dot{y}_n^2}$ is the velocity on entering the gap, $r=v_n/\omega_0$, and ϕ is the angle between the velocity vector v_n and the x axis on entry to the gap.

We supplement the system (3) with an equation for the point where the ion meets the opposite boundary of the modulator gap

$$[x(t)-x_1]\sin\Theta-[y(t)-y_1]\cos\Theta=0, \tag{5}$$

where (x_1,y_1) are the coordinates of an arbitrary point on the opposite boundary of the modulator gap.

Eliminating $x(t)$ and $y(t)$ from Eqs. (3) and (5), we obtain an equation for t which is the time when the ion reaches the opposite boundary of the modulator gap

$$\begin{aligned} r\cos[\omega_0(t-t_n)-\phi+\Theta] &= h_{C1}+D\left\{\sin\omega t-\frac{\omega}{\omega_0}\cos\omega t_n\sin[\omega_0(t-t_n)]\right. \\ &\left.-\sin\omega t_n\cos[\omega_0(t-t_n)]\right\}, \end{aligned} \tag{6}$$

where $h_{C1}=(x_C-x_1)\sin\Theta-(y_C-y_1)\cos\Theta$ is the distance between the center of the circle along which the ion would propagate in a uniform magnetic field and the boundary of the modulator gap.

The transcendental equation (6) can be solved numerically, with the first term in braces playing the role of a correction to the time t_0 which is calculated from the equation for zero electric field and $D=0$,

$$\omega_0(t_0-t_n)=\phi-\Theta+\arccos(h_{C1}/r). \tag{7}$$

Note that Eq. (7) gives an exact solution for the problem of determining the point of intersection of the ion trajectory in a uniform magnetic field with a given planar boundary, for example, with the slit planes of the mass analyzer. Then for the modulator gap, the method of solution involves the normal iteration procedure

$$\begin{aligned} \omega_0(t_{i+1}-t_n) &= \phi-\Theta+\arccos(h_{C1}/r) \\ &+\frac{D}{r}\left\{\sin\omega t_i-\frac{\omega}{\omega_0}\cos\omega t_n\sin\right. \\ &\left.\times[\omega_0(t_i-t_n)]-\sin\omega t_n\cos[\omega_0(t_i-t_n)]\right\}, \end{aligned} \tag{8}$$

which converges very rapidly and gives a relative calculation accuracy of 10^{-5} for three or four iterations. Substituting the value obtained for the time t into Eq. (3), we obtain the ion coordinates at the exit boundary of the modulator gap. The ion velocity components are then determined from equations obtained from Eq. (3) by a single differentiation with respect to time

$$\begin{aligned} \dot{x}(t) &= v_n\cos[\omega_0(t-t_n)-\phi] \\ &+D\{\omega\cos\omega t_n\sin[\omega_0(t-t_n)-\Theta] \\ &-\omega_0\cos\Theta\sin\omega t+\omega_0\cos[\omega_0(t-t_n)-\Theta] \\ &\times\sin\omega t_n+\omega\sin\Theta\cos\omega t\}, \end{aligned}$$

$$\begin{aligned} \dot{y}(t) &= -v_n\sin[\omega_0(t-t_n)-\phi] \\ &+D\{\omega\cos\omega t_n\cos[\omega_0(t-t_n)-\Theta] \\ &-\omega_0\sin\Theta\sin\omega t-\omega_0\sin[\omega_0(t-t_n) \\ &-\Theta]\sin\omega t_n-\omega\cos\Theta\cos\omega t\}. \end{aligned} \tag{9}$$

The initial parameters are set as: the ion coordinates at the exit slit of the ion source $(x_{n0},y_{n0}\equiv 0)$, the ion energy and angle of emission from the source $[qU_0$, $\alpha=\arctan(\dot{x}_{n0}/\dot{y}_{n0})]$, and the phase of the rf modulator voltage ωt_{n0}. In addition, the following parameters are defined to simulate the mass analyzer: the magnetic field H_0, the amplitude of the rf modulating voltage U, the charge and mass of the ions being studied m/q, the geometric parameters of the analyzer, i.e., the diameters of the working orbits D_0, D_1, and D_2, which determine the coordinates of the centers of the analyzer slits, the widths of the ion source slit S_0, the modulator slit S_1, the drift slit S_2, the analyzer exit slit S_3, and the aperture slit S_{A0}, located on the orbit at 90° from the source slit, and the geometric parameters of the modulator d_1, d_2, and d_3.

A program was written for an IBM PC in the languages FoxPro and C. The calculation part of the program gives the results in numerical and graphical form. The other part of the program provides a convenient interface for entering the analyzer parameters, semiautomatic tuning of the parameters of the modulating device to the operating regimes, and facilities for collecting the results in a database for further processing.

It was shown in Refs. 2, 10, 12, and 13 that the operating regime of the MRMS analyzer is subject to certain additional relationships between the defined parameters of the mass analyzer. Finding the set of parameters for which the mass peak is observed is equivalent to tuning a real device. The program tuning regime is based on calculating the central ion trajectory and estimating the first-order aberrations which can be used to obtain the direct ejection voltage, to estimate the precision of tuning to (for example) a compensation mode of operation, the resolution, and the efficiency of utilization of the current from the ion source.

During checks on the program and at the initial stages of the calculations we investigated how the final results of the calculations were influenced by the number of steps over the source exit slit, the angles of emission of the ions from the exit slit, and the ion energy and frequency of the modulating voltage. An increase in the number of parameter steps between a factor of 2 and a factor of 10 significantly increased the computation time although the resolution and other characteristics varied negligibly.

3. The following factors associated with the design characteristics and operation of the device were taken into account in the calculations whose results are presented below. The precision electromagnet developed specially for this MRMS has pole pieces of diameter 600 mm and an interpole gap of width 53 mm. When designing the magnet, we assumed that the working orbit of the MRMS would be located at a distance of approximately two interpole gaps from the

edges of the pole pieces so that edge effects had a minimal influence on the uniformity of the magnetic field for a working orbit of diameter D_1. Thus, only those variants in which D_1 was 400 ± 10 mm were considered in the calculations. Preliminary calculations showed that the nonuniformity of the field on the working orbit $\Delta H/H_0$ (where H_0 is the field strength on the symmetry axis of the pole pieces at the center of the gap) does not exceed $\pm7\times10^{-6}$ with no additional shimming. The design and characteristics of the electromagnet will be described in detail after the final setting up and alignment.

Since the MRMS ion source is situated inside the analyzer chamber, its geometric dimensions should be small. For the device under construction we developed a small ion source whose ion-optical system incorporated a crossed-slit three-electrode lens which could focus the ion beam in the horizontal and vertical planes.[14] The accelerating potential difference U_0 in the source was $\sim$2000 V, so that instabilities of the accelerating voltage, scattered electric fields, contact potential differences, and space charge had a negligible influence on the beam propagation over the initial orbit of diameter D_0 and did not reduce the resolution of the device. From this point of view it would also be advantageous to increase U_0 substantially (several times or even tens of times). However such an increase would involve an almost proportional increase in the geometric dimensions of the insulating elements and the entire ion source, as well as the amplitude of the modulating rf sinusoidal voltage $U_{\rm rf}$, which would be extremely difficult to stabilize. Thus, the calculations were made for the range $U_{\rm rf}$ = 300–600 V in which the required stability of $U_{\rm rf}$ is reasonably attainable.

Since changes in the orbit parameters of the analyzer lead to changes in the direct ejection amplitude U_{13}, the condition $U_{\rm rf}\leqslant0.9U_{13}$ was used to obtain a unique treatment of the calculated results, where $U_{\rm rf}$ is the amplitude of the modulating voltage required to achieve the compensation mode of the MRMS. Another constraint taken into account in the calculations was imposed by the dielectric strength of the accelerating gaps in the modulator. As a result, the gap widths d_1 and d_3 were set at greater than 0.5 mm.

4. One of the most important aspects from the point of view of the design of the device is to determine how the resolution R and the current efficiency K_I are influenced by the diameters D_2 and D_0 and by their difference D_2-D_0.

The diameter difference D_2-D_0 determines the amplitude of the rf modulating voltage needed to ensure that ions of a particular mass and charge M/q leaving the source, enter the analyzer exit slit S_3 after the first pass through the modulator, i.e., in the direct ejection regime. The direct ejection peak corresponds to the lowest amplitude of the rf voltage U_{13} at which the ions enter S_3. Therefore, the value of U_{13} uniquely relates the geometric D_0, D_2 and electrical U_0, ω parameters of the motion of ions having a particular ratio M/q in the magnetic field H and can thus be used as a relative measure of the modulating voltage $U_{\rm rf}$. In most of the calculations we used the relation $U_{\rm rf}=0.9U_{13}$ since first, the absolute value of $0.1U_{13}$ is 30–50 V, which prevents nonresonant ions from entering the exit slit, and second as $U_{\rm rf}$ increases, the resolution of the MRMS increases, i.e., it is best to operate at high modulating voltages.

TABLE I.

D_0, cm	D_2, cm	D_2-D_0, cm	$U_{\rm rf}\approx0.9U_{13}$	$R_{50\%}/10^6$	K_I
36.0	42.2	6.2	351	1.076	0.0078
36.2	42.2	6.2	351	1.075	0.0078
36.4	42.6	6.2	352	1.076	0.0078
36.6	42.8	6.2	353	1.078	0.0077
36.8	43.0	6.2	353	1.077	0.0077
37.0	43.2	6.2	353	1.074	0.0077
36.0	43.2	7.2	409	1.224	0.0068
36.2	43.4	7.2	410	1.225	0.0068
36.4	43.6	7.2	411	1.228	0.0067
36.6	43.8	7.2	411	1.225	0.0067
36.8	44.0	7.2	411	1.223	0.0067
37.0	44.2	7.2	411	1.220	0.0067

Table I gives the results of calculations of $R_{50\%}$ and K_I as a function of D_0 and D_2 for two different values of the difference (D_2-D_0). In this case, the diameter of the working orbit D_1 is calculated assuming that the MRMS is operating in the compensation mode.[12,13] For these differences D_2-D_0 of 6.2 and 7.2 cm the required values of $U_{\rm rf}=0.9U_{13}$ are 352 ± 1 and 410 ± 1 V, respectively. The calculations show that as the diameter difference D_2-D_0 increases, both $U_{\rm rf}$ and the resolution increase. The values of the diameters D_0 and D_2 do not influence $R_{50\%}$. For its part the current efficiency K_I, which determines its sensitivity, decreases with increasing D_2-D_0 so that the product $R_{50\%}K_I$ remains almost constant.

Figure 2 gives $R_{50\%}$ and K_I as a function of the modulating voltage $U_{\rm rf}$ for two values of the diameters D_0 and D_2 and also their differences D_2-D_0. The calculations show that for each set of geometric parameters D_0, D_2, and D_2-D_1 the resolution $R_{50\%}$ increases while the current efficiency K_I decreases with increasing $U_{\rm rf}$. This improvement in the resolution with increasing $U_{\rm rf}$ can be explained, first, by the fact that the dispersion of the device increases and, second, the deflection of the packets cut out by the drift slit S_2 is reduced, which indicates a fall in the modulation aber-

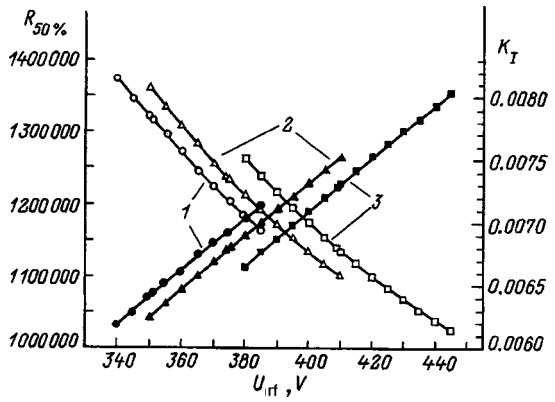

FIG. 2. Resolution $R_{50\%}$ and current efficiency versus amplitude of modulating voltage: $(D_2-D_0)1=6.2$ (*1*), 6.6 (*2*), 7.2 cm (*3*); ○,△,□ — K_I, ●,▲,■ — $R_{50\%}$.

rations. These two factors also explain why an increase in $R_{50\%}$ is observed with increasing difference D_2-D_0 for relative values of $U_{\mathrm{rf}}=0.9U_{13}$.

The decrease in K_I is caused by an increase in the modulation amplitude of the ion beam and a reduction in the number of ions in packets cut out by the drift slit S_2. These calculated dependences $R_{50\%}=f(U_{\mathrm{rf}})$ and $K_I=f(U_{\mathrm{rf}})$ show good agreement with the experimental data obtained using the MRMS prototype.[13]

The calculated data given in Table I and plotted in Fig. 2 show that in order to achieve a high resolution, it is best to have large differences D_2-D_0 and high values of U_{rf}.

It is known from the theory that in any type of mass spectrometer the resolution is directly proportional to the dispersion and inversely proportional to the width of the ion beam image on the plane where the analyzer exit slit is located. The width of the image includes the width of the slit in the ion source (in our case S_0), the exit slit (S_3), and all the aberrations which contribute to the ion beam broadening. In the calculations we allowed for the aberrations caused by the spread of ions over angles of emission from the source and energies, and also beam aberrations caused by the modulation process.

In order to calculate the magnetic aberrations, we need to know the real distribution of magnetic field inhomogeneities on the drift orbit of diameter D_1 which is not known exactly at present. Thus, on the basis of investigations carried out using the MRMS prototype we estimated that the aberrations caused by these inhomogeneities may be between 5 and 15 μm. Bearing in mind that in preliminary investigations the magnitude of the magnetic field inhomogeneities ΔH on the drift orbit did not exceed $\pm 7\times 10^{-6}$ of H_0, these estimates of the magnetic aberrations seem realistic. Consequently, the contribution of these aberrations to reducing the resolution of the device does not exceed $\sim$10%.

We shall now analyze how the widths of all the slits determining the ion trajectory from the ion source to the detector influence the resolution and the current efficiency of the MRMS.

Since the width of the ion-source exit slit S_0 (when varied over a small range close to the real values) barely influences the resolution at the peak half-height $R_{50\%}$, the program was designed to calculate the resolution at 1% peak height by two methods: analytically, using the dispersion and the width of the ion beam image, and also from the peak profile. Graphs of $R_{1\%}\sim f(S_0)$ obtained by the two methods are plotted in Fig. 3. The substantial difference between the values of $R_{1\%}$ is evidently attributable to the increased influence of the ''tails'' at the base of the line in the analytical method.

The sensitivity of the MRMS is determined not only by the width of the ion-source exit slit S_0 but also by the focusing properties of the specially developed ion sources[14,15] and by the current efficiency K_I which is equal to the ratio of the current entering the exit slit of the analyzer S_3 to the current leaving S_0. Assuming that the slit S_0 is uniformly filled with ions, reducing the width S_0 will lead to a proportional reduction in the output current of the source. However, by improving the focusing properties of the ion optical system, including vertical focusing of the ion beam,[15] the output current from the source and the overall sensitivity of the MRMS prototype were improved by almost an order of magnitude with the resolution being maintained. Although in most of the calculations the slit width S_0 was taken to be 28 μm, we considered it quite permissible to reduce S_0 to 20–25 μm.

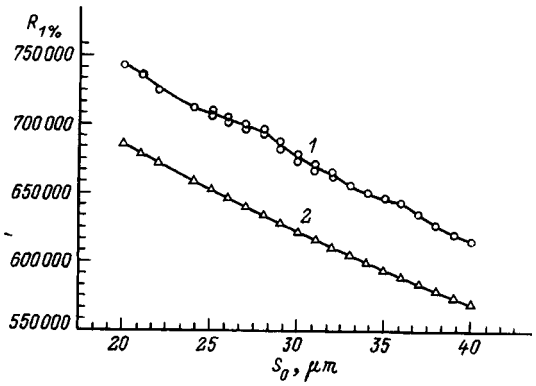

FIG. 3. Resolution $R_{1\%}$ calculated by two methods as a function of the width of the source exit slit.

The width of the aperture slit S_{A1} positioned at 90° from the source determines the angular aberrations of the ion beam and can therefore influence the resolution and sensitivity of the MRMS. However, since the angular aberrations make a small contribution to the beam image ($\leqslant 1$ μm) if the total width of the image is $\sim$50 μm, varying S_{A1} between 0.2 and 2.0 mm leads to changes in $R_{50\%}$ of 1–2%. The final slit width S_{A1} will be chosen experimentally when the ion source is tuned.

The three slits in the modulator electrodes, having the same width and denoted as S_1, do not directly determine the resolution and sensitivity of the MRMS since their width is tens of times greater than the width of the source exit slit S_0. However, it is quite clear that for any slit widths S_1 in the modulator, there is an electric field sag which indirectly influences the modulator efficiency and the resolution of the device. Investigations using the MRMS prototype[13] showed that the presence of these sagging electric fields is similar to increasing the width of the accelerating gaps of the modulator d_1 and d_3, which lowers the resolution of the device. This effect can be compensated by increasing the amplitude U_{rf}.

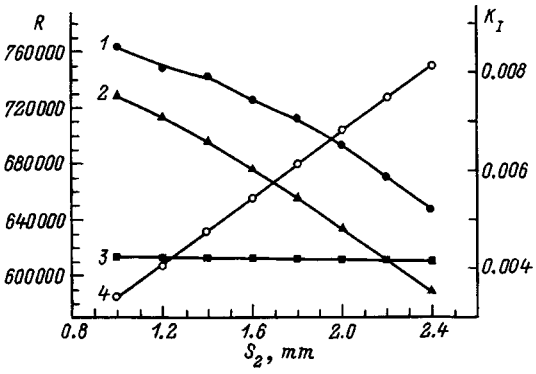

FIG. 4. Resolution $R_{1\%}$ and current efficiency versus width of drift slit: *1* — $R_{1\%}$, *2* — R_{gen}, *3* — $R_{50\%}$, and *4* — K_I.

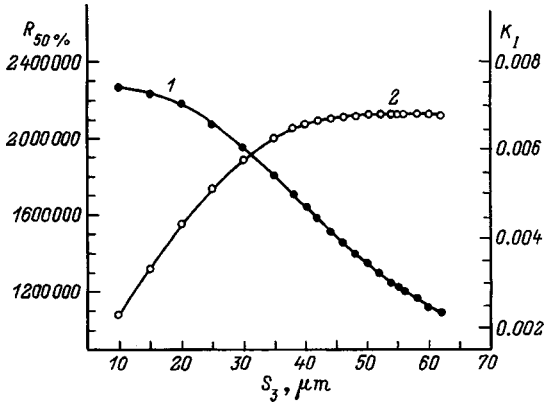

FIG. 5. Dependences of $R_{50\%}$ (*1*) and K_I (*2*) on width of analyzer exit slit.

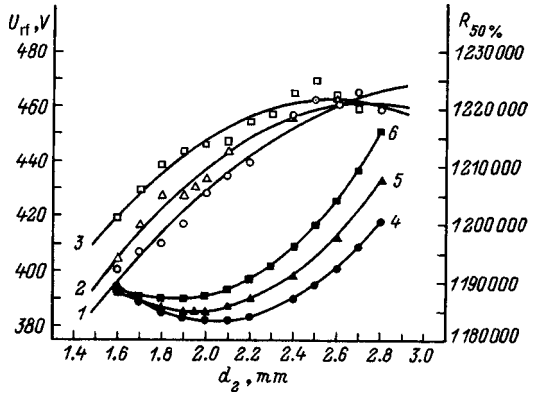

FIG. 7. Dependences of $R_{50\%}$ and U_{rf} on the width of the field-free modulator space d_2; $d_1 \neq d_3$ mm: *1, 4* — 0.8, *2, 5* — 0.9, *3, 6* — 1.0, *1–3* — $R_{50\%}$, and *4,6* — U_{rf}.

However, for selected widths of the accelerating (d_1, d_3) and field-free (d_2) gaps in the modulator, sagging of the fields in the electrode slits imposes constraints on the ion energy spread and the higher U_{rf}, the more stringent are these constraints.[13] Bearing in mind that in the MRMS prototype, the amplitude U_{rf} was ~570 V but in the device being developed, this is ~400 V, the influence of the sagging electric fields should be weaker. Thus, in the calculations the slit widths S_1 were taken to be the same as in the prototype, i.e., $S_1 \approx 1$ mm.

The drift slit S_2 sets the diameter of the working orbit D_1 and cuts out ion packets from the modulated ion beam. Changing the width S_2 causes a proportional change in the current efficiency K_I and an inversely proportional change in the resolution $R_{1\%}$. This is because as S_2 decreases, the aberration caused by modulation of the beam decreases. The calculations show that when S_2 is increased from 1.0 to 2.4 mm, the modulation aberrations increase from 4.5 to 26 μm. Figure 4 gives the dependences $R_{1\%} \approx f(S_2)$ and $K_I = f(S_2)$. The optimum width of the drift slit S_2 is in the range 1.7–2.0 mm for which the modulation aberrations are 11–18 μm.

Figure 5 gives the resolution $R_{50\%}$ and the current efficiency K_I as a function of the width of the analyzer exit slit S_3 for $S_0 = 28$ μm, $S_2 = 2.0$ mm, and $U_{rf} = 0.9U_{13}$.

The optimum slit width S_3 to maximize K_I is equal to the width of the ion beam image allowing for all aberrations. It can be seen from the graph that the optimum slit width S_3 is ~50 μm and in the range 40–50 μm the relative drop in K_I is appreciably less than the relative increase in $R_{50\%}$. In order to compensate for magnetic aberrations neglected in the calculations, the value of S_3 should be increased toward the upper limit. Thus, in the final calculations the width of the analyzer exit slit S_3 was varied between 45 and 50 μm.

The most difficult results to interpret are those of calculations to determine how the resolution of the MRMS is influenced by the widths of the accelerating gaps in the modulator d_1 and d_3, and also by the width of the field-free space in the central electrode d_2. One the one hand, if the diameters D_0 and D_2 are kept constant, changes in the widths d_1, d_2, and d_3 lead to a change in the diameter of the working orbit D_1. This must be accompanied by a change in the magnitudes and phase of the modulating voltage U_{rf} to ensure that the MRMS operates in the compensation mode. On the other hand, if the widths of the accelerating gaps d_1 and d_3 are varied, the electric field in these gaps and thus the ion trajectories vary (trochoids). The result of the influence of these two factors is that the calculated resolution of the MRMS remains almost constant in the presence of appreciable changes in d_1, d_2, and d_3 if the condition

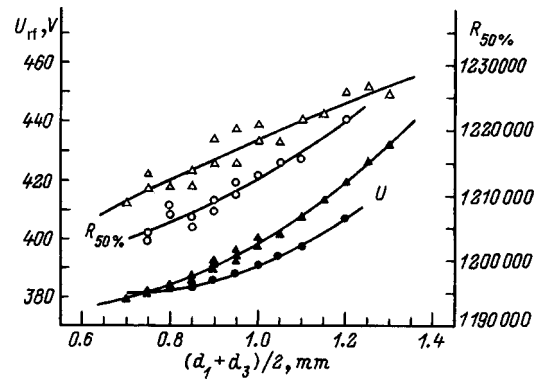

FIG. 6. Resolution $R_{50\%}$ versus width of accelerating gaps d_1 and d_3 for symmetric ($d_1 = d_3$) and asymmetric gaps ($d_1 \neq d_3$): modulator $d_2 = 2.0$ (circles) and 2.2 mm (triangles).

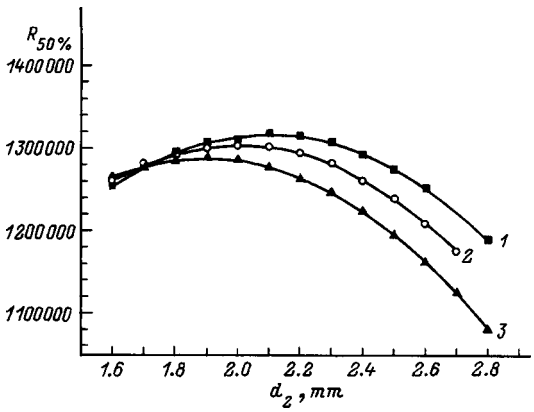

FIG. 8. Dependences of $R_{50\%}$ on the width of the field-free space d_1, d_2, d_3 for $U_{rf} \neq 0.9U_{13}$, $d_1 = d_3$, mm: *1* — 0.8, *2* — 0.9, and *3* — 1.0.

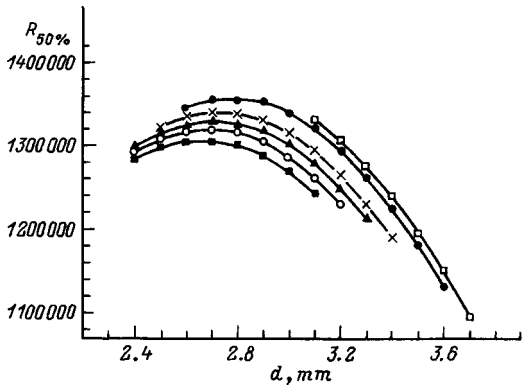

FIG. 9. Dependences of $R_{50\%}$ on the width of the modulator gaps d_1, d_2, d_3 for $U_{rf} \neq 0.9U_{13}$; $d=(d_1/2)+d_2(d_3/2)$; d_2, mm: ■ — 1.9, ◇ — 2.0, ▲ — 2.1, × — 2.2, ● — 2.4, and ○ — 2.5.

$U_{rf}=0.9U_{13}$ is maintained (Figs. 6 and 7). In this case, U_{rf} has a minimum in the range $d_2 \approx$ 1.9–2.1 mm.

If the condition $U_{rf}=0.9U_{13}$ is not satisfied (e.g., if U_{rf} remains constant while the parameters d_1, d_2, and d_3 vary), the family of curves $R_{50\%}=f(d_1,d_2,d_3)$ has a maximum in the range $d_2 \approx 1.82$–2.2 mm (Fig. 8) and $(d_1/2+d_2+d_1/2) \approx 2.7$–2.9 mm (Fig. 9) both for symmetric accelerating gaps $d_1=d_3$, and for asymmetric gaps $d_1 \neq d_3$. From this it follows that the width of the accelerating gaps is in the range 0.7–0.9 mm. The minimum values of U_{rf} specifically correspond to these values of d_1 and d_3 (Fig. 7).

The final calculations of the analytical characteristics ($R_{50\%}$, $R_{1\%}$, and K_I) of the MRMS analyzer were made with the parameters varied simultaneously in the narrow limits specified at the preliminary stages. As a result, we determined the optimum analyzer parameters for which the rated resolution is $R_{50\%} \approx 1.35 \times 10^6$ with a current efficiency $K_I \approx 0.006$.

In the course of the calculations we identified correlations which can be used, for example, to compensate for the influence of some parameters on $R_{50\%}$ by varying another parameter, which is required to tune and investigate the MRMS being built.

The calculations also showed that this MRMS system is stable over a wide range of geometric, electrical, and frequency parameters. Comparing the results of the calculations used to construct a MRMS with $R_{50\%} \approx 350\,000$ (Refs. 11 and 12) with the results obtained here, we note some significant advantages (such as an improvement in the resolution of the static stage of the device by increasing the orbit radius, an appreciable reduction in the amplitude of the rf modulating voltage U_{rf}, and an increase in the diameter differences D_1-D_0 and D_2-D_1), which in our opinion indicate that the calculated analytical characteristics can be achieved in practice.

This work was supported by a grant and carried out under the auspices of the State Scientific and Technical Program ''Fundamental Metrology.''

[1] N. I. Ionov, B. A. Mamyrin, and V. B. Fiks, Zh. Tekh. Fiz. **23**, 2101 (1953).
[2] B. N Shustrov, Zh. Tekh. Fiz. **30**, 860 (1960) [Sov. Phys. Tech. Phys. **5**, 810 (1960)].
[3] B. A. Mamyrin and A. A. Frantsuzov, Prib. Tekh. Éksp. **3**, 114 (1961).
[4] Alekseĭchuk, G. S. Anufriev, G. I. Afonina *et al.*, Prib. Tekh. Éksp. **4**, 206 (1979).
[5] G. S. Anufriev, L. A. Arbuzova, B. S. Boltenkov *et al.*, Zh. Anal. Khim. **38**, 472 (1983).
[6] I. N. Tolstikhin, I. Ya. Azbel', and L. V. Khabarin, Geokhimiya **5**, 653 (1975).
[7] G. S. Anufriev, B. S. Boltenkov, and I. N. Kapitonov, Dokl. Akad. Nauk **304**, 702 (1989).
[8] Yu. A. Akulov, B. A. Mamyrin, L. V. Khabarin *et al.*, Pis'ma Zh. Tekh. Fiz. **14**, 940 (1988) [Sov. Tech. Phys. Lett. **14**, 416 (1988)].
[9] B. A. Mamyrin and A. A. Frantsuzov, Zh. Éksp. Teor. Fiz. **48**(2), 416 (1965) [Sov. Phys. JETP **21**, 274 (1965)].
[10] B. A. Mamyrin, N. N. Aruev, and S. A. Alekseenko, Zh. Éksp. Teor. Fiz. **63**, 3 (1973) [Sov. Phys. JETP **36**, 1 (1973)].
[11] B. A. Mamyrin, N. N. Aruevr, and S. A. Alekseenko, Zh. Éksp. Teor. Fiz. **80**, 2125 (1981) [Sov. Phys. JETP **53**, 1109 (1981)].
[12] N. N. Aruev, E. L. Baĭdakov, B. A. Mamyrin, and A. V. Yakovlev, Zh. Tekh. Fiz. **57**, 502 (1987) [Sov. Phys. Tech. Phys. **32**, 303 (1987)].
[13] N. N. Aruev, E. L. Baĭdakov, B. A. Mamyrin, and A. V. Yakovlev, Zh. Tekh. Fiz. **60**(8), 181 (1990) [Sov. Phys. Tech. Phys. **35**, 980 (1990)].
[14] N. N. Aruev and E. L. Baĭdakov, Zh. Tekh. Fiz. **65**(4), 175 (1995) [Tech. Phys. **40**, 385 (1995)].
[15] N. N. Aruev, E. L. Baĭdakov, B. A. Mamyrin, and A. V. Yakovlev, Zh. Tekh. Fiz. **58**, 536 (1988) [Sov. Phys. Tech. Phys. **33**, 321 (1988)].

Translated by R. M. Durham

Determination of the thermal conductivity of polycrystalline diamond films by means of the photoacoustic effect

A. N. Obraztsov and I. Yu. Pavlovskiĭ

M. V. Lomonosov State University, 119899 Moscow, Russia

V. G. Ral'chenko

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

H. Okushi and H. Watanabe

Electrotechnical Laboratory, Tsukuba, Ibaraki 305, Japan

(Submitted September 30, 1997; resubmitted March 23, 1998)
Zh. Tekh. Fiz. **69**, 97–101 (April 1999)

A new method of determining the heat-conducting properties of diamond films is proposed, based on the photoacoustic effect. This method is used to study diamond polycrystalline films grown on silicon by chemical vapor deposition in a microwave discharge plasma. The thermal conductivity obtained was approximately half that for single-crystal diamond. © *1999 American Institute of Physics.* [S1063-7842(99)01704-3]

INTRODUCTION

Among the many remarkable properties of diamond, one of the most interesting from the applied and fundamental point of view is its thermal conductivity. We know that diamond single crystals have the highest thermal conductivity of known materials (up to $25\,\mathrm{W\cdot cm^{-1}\cdot K^{-1}}$ at room temperature[1]), which, in particular, makes diamond a promising material for fabricating efficient heat transfer systems for microelectronics devices such as laser and microwave diodes, multichip modules, and so on in which a large quantity of heat is released locally. High thermal conductivity may prove one of the advantages of fabricating various microelectronics devices using doped diamonds with semiconducting properties. The extensive prospects for the practical use of diamonds (including as heat transfer systems) are to a large measure attributable to the development of techniques for synthesizing diamond films by chemical vapor deposition (CVD).[2]

Despite the major interest in this problem, no reliable methods of measuring the thermal conductivity of CVD diamond films have yet been developed. Usually, the thermal diffusion length is measured experimentally and the thermal conductivity is calculated using tabular data on the density and specific heat of diamond.[3–6] The fact that these parameters for polycrystalline films may deviate from the tabular values and also the spatial nonuniformity in the properties of the films and other factors lead to errors in calculating the thermal conductivity. The thermal conductivity can be determined directly by methods based on measuring the steady-state temperature distribution in a sample with a particular configuration.[7,8] However, the high thermal conductivity of diamond materials makes these measurements fairly complex and in addition, specially shaped film samples are required, which limits the range of application of these methods. In particular, it is impossible to make local measurements of the thermal conductivity and nondestructive investigations of large-area films. A consequence of these problems is that measurements made in different laboratories give completely different values of the thermal conductivity, even for identical film samples.[9] There is thus a need to search for new methods of measuring the thermal conductivity which would be effective for studying diamond materials.

PHOTOACOUSTIC EFFECT IN DIAMOND FILMS

In the present paper we analyze the possibility of measuring the thermal conductivity of diamond films using a method based on the photoacoustic effect, whose physical nature is similar to the mirage effect widely used to study thermal conductivity (including that in diamond films[5,9]). In the mirage effect, a laser probe beam propagating along the surface of an object is displaced as a result of refraction in a gas or liquid in contact with the surface and heated nonuniformly by the heat released when the object absorbs light from another fairly high-power light source (pulsed laser). One usually measures the time delay of the displacement relative to the light pulse (or the phase shift, if the light is periodically modulated).

In the photoacoustic effect, an acoustic wave forms in the gas surrounding the sample as a result of the sample being heated by absorbing an intensity-modulated light beam. The amplitude and phase of the acoustic wave are measured using a sensitive microphone placed together with the sample in a hermetically sealed measuring cell.[10,11] This factor, which distinguishes the photoacoustic effect from the mirage effect, means that the thermal conductivity can be obtained directly from the experimental data without using additional parameters.

A theoretical analysis of the photoacoustic effect yields a general expression for the acoustic wave amplitude[10,11] al-

though this is fairly cumbersome and difficult to interpret. In some cases of practical importance this expression can be simplified substantially. For example, if the sample thickness is substantially greater than the effective light absorption depth $1/\beta$ (where β is the absorption coefficient), which in turn is smaller than the characteristic thermal diffusion length $\mu=(k/\pi\rho C\nu)^{1/2}$, where k is the thermal conductivity, ρ is the density, and C is the specific heat of the material being studied, the acoustic wave amplitude q can be given as

$$q=Y_1\frac{1}{\nu\sqrt{k\rho C}}. \tag{1}$$

If the light penetration depth is greater than the thermal diffusion length $(1/\beta>\mu)$, then

$$q=Y_2\frac{\beta}{\rho C\nu^{3/2}}, \tag{2}$$

where Y_1 and Y_2 are constants comprising a combination of parameters determined by the experimental conditions, the properties of the gas filling the measuring cell, and the base material on which the sample is deposited.

It is easy to see from Eqs. (1) and (2) that if the measurements are made at the same light modulation frequency for two samples of the same shape, the corresponding amplitudes of the photoacoustic signal q_1 and q_2 will be related by

$$\frac{q_1}{q_2}=\sqrt{\frac{k_2\rho_2C_2}{k_1\rho_1C_1}} \tag{3}$$

if the light absorption depth is less than the thermal diffusion length, and by

$$\frac{q_1}{q_2}=\sqrt{\frac{\beta_1\rho_2C_2}{\beta_2\rho_1C_1}} \tag{4}$$

if the thermal diffusion length μ is less than $1/\beta$.

It can be seen from formulas (3) and (4) that by making measurements at the same frequency ν in different spectral ranges for which the light absorption depth is less than or greater than the thermal diffusion length (the wavelengths of monochromatic light λ_1 and λ_2, respectively), we obtain

$$\frac{k_1}{k_2}=\left(\frac{\beta_2}{\beta_1}\right)_{\lambda_2}\left(\frac{q_1}{q_2}\right)_{\lambda_2}\left(\frac{q_2}{q_1}\right)^2_{\lambda_1}, \tag{5}$$

where the subscripts λ_1 and λ_2 indicate the wavelength at which the quantity in parentheses is measured.

Formula (5) can be used to determine the relative change in the thermal conductivity in different sections of a diamond film or different samples with similar dimensions and ratios of thermal diffusion length to light absorption depth. The absolute value of the thermal conductivity k_1 can be determined if we have a sample with known heat-conducting properties k_2, such as a diamond single crystal. In addition to the amplitude of the photoacoustic signal measured in two spectral ranges, we also need to know the light absorption coefficient β_2 and β_1 in the relatively weak absorption range $(\beta>1/\mu)$. Assuming that over a fairly wide range, the amplitude of the photoacoustic signal is linearly proportional to the light absorption coefficient $(q\sim\beta)$ (Refs. 8–10), formula (5) may be simplified substantially

$$\frac{k_1}{k_2}=\left(\frac{q_1}{q_2}\right)^2_{\lambda_1}. \tag{6}$$

Note that formula (6) was obtained assuming that the samples have fairly similar geometric dimensions (thickness) and similar optical and heat-conducting properties. In addition, the light modulation frequency should satisfy

$$\mu=\sqrt{\frac{k}{\pi\rho C\nu}}<l, \tag{7}$$

where l is the sample thickness.

Substituting tabular data for single-crystal diamond into formula (7), which are clearly the limiting values for polycrystalline films[1] ($k=25\,\mathrm{W\cdot cm^{-1}\cdot K^{-1}}$, $\rho=3.5\,\mathrm{g\cdot cm^{-3}}$, and $C=6.19\,\mathrm{J\cdot g^{-1}\cdot K^{-1}}$), we find that for a 100 μm thick diamond wafer condition (7) will be satisfied for a modulation frequency above 3.5 kHz and that for a 500 μm thick wafer similar to those used in heat transfer systems, the modulation frequency should be less than 150 Hz. This estimate shows that the proposed method can be implemented in relatively simple experiments at modulation frequencies attainable with normal mechanical choppers.

EXPERIMENTAL VERIFICATION OF THE METHOD

The proposed method of measuring the thermal conductivity was checked out experimentally using 4×4 mm, 300 μm thick single crystals of type 1b synthetic diamond and also samples of polycrystalline films grown in a microwave discharge plasma as described by Graebner *et al.*[6] A diamond film 57 μm in diameter was grown on a silicon substrate and then separated by etching the silicon in a mixture of hydrofluoric and nitric acids. The film thickness varied between $\sim$250 and $\sim$200 μm from the center toward the edge. Six 8×8 mm samples were cut from different sections along the radius of this disk using a focused laser beam.

The prepared samples exhibited morphology typical of diamond polycrystalline films with characteristic clearly defined faceting of randomly oriented crystallites. The Raman spectra of the films recorded using 633 nm helium–neon laser radiation revealed a single line at 1322 cm^1 with a full width at half-maximum of around 4 $\mathrm{cm^{-1}}$, which corresponds to crystalline diamond (Fig. 1). The almost complete absence of any light scattering in the spectral range 1350–1800 $\mathrm{cm^{-1}}$ indicates that the films contain no nondiamond carbon.[12] A characteristic feature of these films was the observation of intense luminescence centered around 740 nm (Fig. 2) (excited using the 633 nm helium–neon laser line), which indicates that the diamond crystal lattice contains silicon inclusions.[13,14]

The photoacoustic measurements were made using an updated photoacoustic spectrometer made by Princeton Applied Research Corporation (Model 6001). The light source was a 1 kW xenon lamp. A grating monochromator was used to select light in a particular spectral range. The amplitude of

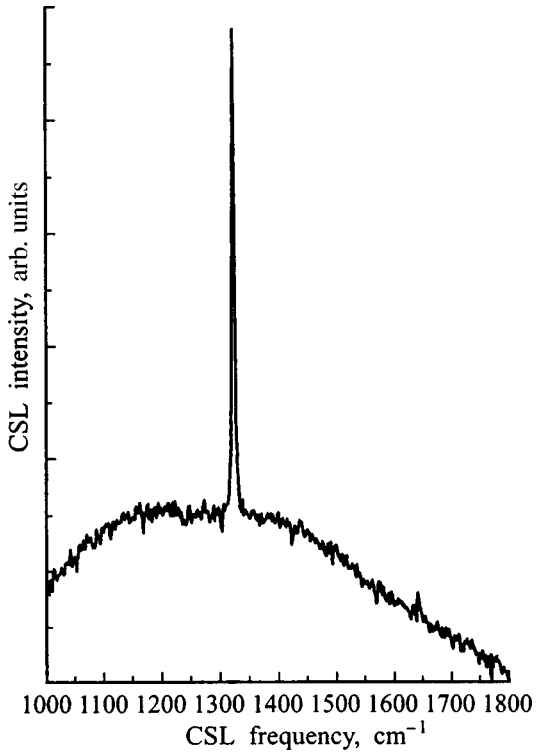

FIG. 1. Raman spectrum of CVD diamond film.

the photoacoustic signal was normalized to the flashlamp spectrum by means of a pyroelectric detector. The light was modulated by a mechanical chopper at frequencies between $\nu = 20$ Hz and 5 kHz. The size of the light spot on the sample was 3×3 mm. All the measurements were made at room temperature in an air-filled cell.

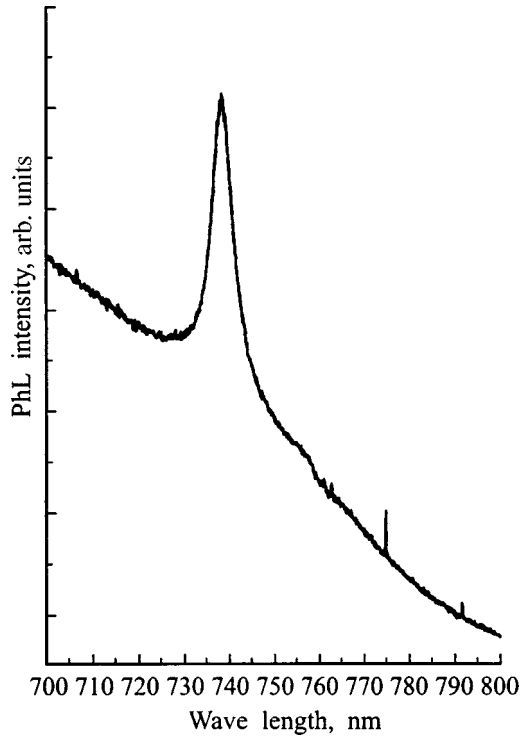

FIG. 2. Photoluminescence spectrum of CVD diamond film.

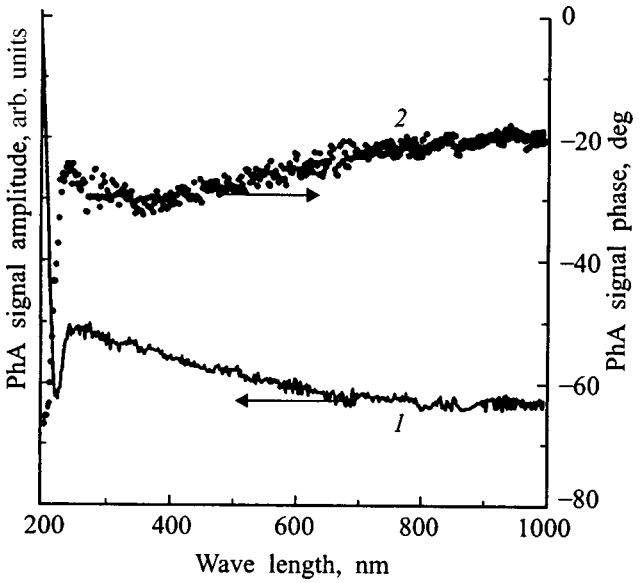

FIG. 3. Spectral dependences for the amplitude (*1*) and phase (*2*) of the photoacoustic signal.

Figure 3 shows typical spectral dependences for the amplitude and phase of the photoacoustic signal recorded at a modulation frequency $\nu = 90$ Hz. Since the photoacoustic amplitude is proportional to the absorption coefficient over a wide range of values,[10,11] the amplitude spectra in Fig. 3 represent the absorption spectra of the diamond film. Their shape is typical of diamond polycrystalline films and is indicative of band–band absorption for photons whose energy is higher than the diamond band gap of 5.4 eV (225 nm). In the range 225–250 nm the amplitude of the photoacoustic signal falls appreciably because the fraction of light absorbed in the film is reduced as a result of reflection from the plane faces of the diamond crystallites.[15] The phase of the photoacoustic signal depends on the ratio of the characteristic light absorption and thermal diffusion lengths, and its spectral dependence also reflects the spectral dependence of the light absorption coefficient.

The absence of any clearly defined absorption lines in the visible, which are a consequence of various structural defects and amorphous carbon impurities in the diamond films,[15,16] correlates well with the Raman spectra shown in Fig. 1.

It was established by direct measurements in transmission that the absorption coefficient of light in the range 200–220 nm is at least 10^5 cm^{-1}, which satisfies one of the conditions used to obtain the initial formulas (1) and (2). The second condition ($\mu < l$) is demonstrated clearly in Fig. 4, which gives the frequency dependence of the ratio of the photoacoustic signal amplitudes for a diamond single crystal and samples of diamond films (q_2/q_1) cut from the central part of the 57 mm wafer (curve *1*) and from its edge region (curve *2*). At low modulation frequencies (up to $\nu \sim 200$ Hz) the thermal diffusion length exceeds the thickness of the films and the single crystal, as follows from formula (7) and estimates made earlier, and the amplitude of the photoacoustic signal for both types of samples is determined mainly by the heat-conducting properties of the material used for the

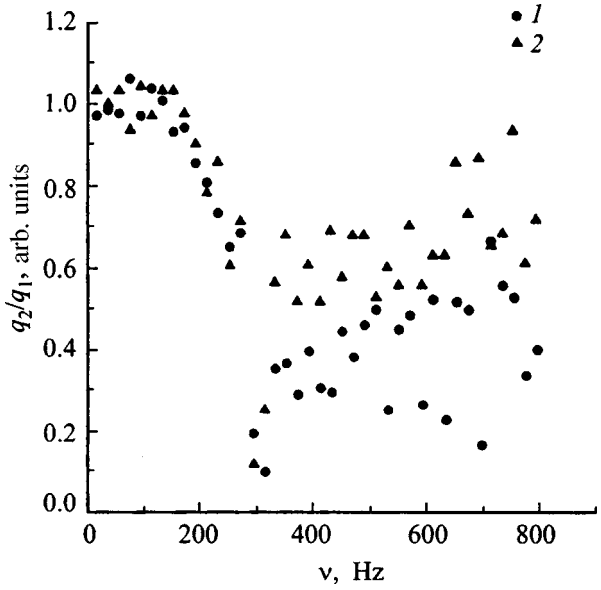

FIG. 4. Frequency dependences of the ratio of the photoacoustic signal amplitudes for two samples of diamond film.

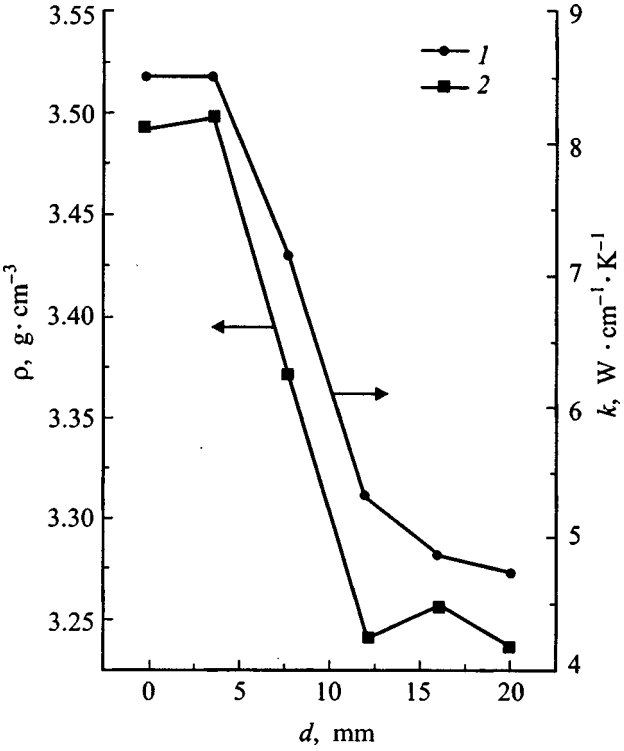

FIG. 5. Thermal conductivity k (*1*) and density ρ (*2*) for different sections of the same diamond film as a function of the distance measured along the radius from the center.

measuring cell (stainless steel).[10,11] Thus, the amplitude ratio q_2/q_1 in this frequency range is close to 1. At higher frequencies the thermal diffusion length becomes appreciably less than the sample thickness, and the ratio q_2/q_1 is determined by the ratio of the thermal conductivities as given by formula (6). The fact that some points on the frequency dependence lie outside the general dependence may be because at the corresponding frequencies ($\sim$300 Hz) the thermal diffusion length in the polycrystalline film is less than its thickness, whereas this condition is not satisfied for the diamond single crystal. Substituting into formula (7) the values $l=250\,\mu$m and $\nu=300$ Hz and the tabular values for diamond, $\rho=3.5\,\mathrm{g\cdot cm^{-3}}$ and $C=6.19\,\mathrm{J^{-1}\cdot K^{-1}}$ (Ref. 1), we find that the thermal conductivity k of polycrystalline films is $\sim 13\,\mathrm{W\cdot cm^{-1}\cdot K^{-1}}$.

This estimate can be refined by using our experimental data and formula (6). It can be seen from Fig. 4 that the ratio q_2/q_1 is $\sim$0.7 for the sample cut from the central part of the film and $\sim$0.5 for the sample cut from the edge, i.e., the thermal conductivity of the polycrystalline film is $\sim$0.5 of that of the single crystal in the first case and $\sim$0.25 in the second. Unfortunately we did not have the technical capability to determine the thermal conductivity for the diamond single crystal used in these experiments. However, its value is most likely in the typical range for these crystals, 15–20 $\mathrm{W\cdot cm^{-1}\cdot K^{-1}}$ (Ref. 7). As a result, we find that the thermal conductivity in the central part of the diamond film may be 7–10 $\mathrm{W\cdot cm^{-1}\cdot K^{-1}}$ while that at the edges is 3.5–5 $\mathrm{W\cdot cm^{-1}\cdot K^{-1}}$.

Bearing in mind data on the correlation between the heat-conducting properties of diamond films and the width of the diamond line in the Raman spectra,[18] our values of the thermal conductivity are fully consistent with the observation that the Raman linewidth 4 $\mathrm{cm^{-1}}$ is substantially greater than that for the single crystals ($\sim 2\,\mathrm{cm^{-1}}$). The radial variation of the heat-conducting properties of the polycrystalline diamond film also correlates with the change in its thickness and density determined from the weight of samples with known dimensions (Fig. 5). The deterioration of the thermal conductivity and reduction in the density at the edges of the diamond film may be attributed to the appearance of pores in the structure which are caused by the nature of the diamond deposition in this zone.

The presence of a fairly large quantity of silicon impurities in these films is evidently a consequence of the extremely high degree of activation of the gaseous medium which is required in order to increase the growth rate of the film but also causes the reduction of silicon from quartz components in the reactor with the appearance of silicon in the gas phase, and its subsequent incorporation in the deposited film.[19] At the same time, the high degree of activation leads to secondary nucleation of the diamond and consequently results in the formation of additional intercrystallite boundaries which impede the propagation of heat.

To conclude, we have shown that in principle, the photoacoustic method can be used to determine the thermal conductivity of diamond materials, including diamond polycrystalline films. With some refinement, the proposed method may be used for rapid diagnostics of this fundamentally important parameter of diamond CVD films. The noncontact optical nature of these measurements means that the spatial distribution of the heat-conducting properties of diamond films with a large surface area can be measured.

The authors are grateful to A. A. Smolin and S. I. Voronin for preparing the samples of diamond films for the investigations. One of the authors would like to sincerely thank

the Matsumae International Foundation (Japan) for a scholarship which made this research possible.

[1] *The Properties of Diamond*, edited by J. E. Field (Academic Press, London, 1990, 674 pp.).
[2] B. V. Spitsyn, in *Handbook of Crystal Growth*, Vol. 3 (Elsevier, North-Holland, 1994, p. 403.).
[3] J. E. Graebner, Diamond Films Technol. **3**, 77 (1993).
[4] E. P. Visser, E. H. Versteegen, and W. J. P. van Enckervort, J. Appl. Phys. **71**, 3238 (1992).
[5] T. R. Anthony, W. F. Banholzer, J. F. Fleisher *et al.*, Phys. Rev. B **4**, 1104 (1990).
[6] J. E. Graebner, V. G. Ralchenko, A. A. Smolin *et al.*, Diamond Relat. Mater. **5**, 643 (1996).
[7] E. Wörner, C. Wild, W. Muller-Sebert *et al.*, Diamond Relat. Mater. **5**, 688 (1996).
[8] J. E. Graebner, J. A. Mucha, L. Seibles, and G. W. Kammlott, J. Appl. Phys. **71**, 3143 (1992).
[9] D. Fournier and K. Plaman, Diamond Relat. Mater. **4**, 809 (1995).
[10] J. C. Rosencwaig, *Photoacoustics and Photoacoustic Spectroscopy* (Wiley, New York, 1980, 309 pp.).
[11] A. Rosencwaig and A. Gersho, J. Appl. Phys. **47**, 64 (1976).
[12] P. K. Bachmann and D. U. Wiechert, Diamond Relat. Mater. **1**, 422 (1992).
[13] A. T. Collins, Diamond Relat. Mater. **1**, 457 (1992).
[14] V. S. Vavilov, A. A. Gippius, A. M. Zaĭtsev *et al.*, Fiz. Tekh. Poluprovodn. **14**, 1078 (1980) [Sov. Phys. Semicond. **14**, 1811 (1980)].
[15] A. N. Obraztsov, H. Okushi, H. Watanabe, and I. Yu. Pavlovsky, Diamond Relat. Mater. **6**, 1124 (1997).
[16] A. N. Obraztsov, T. Izumi, H. Okushi *et al.*, Vestn. Mosk. Univ. Ser. Fiz. Astron., No. 3, 45 (1997).
[17] *Diamond: Electronic Properties and Applications*, edited by L. S. Pan and D. R. Kania (Kluwer, Dordrecht, 1995, p. 370.).
[18] J. E. Graebner, M. E. Reiss, L. Seibles *et al.*, Phys. Rev. B **50**, 3702 (1994).
[19] H. Kiyota, H. Okushi, T. Ando *et al.*, in *Proceedings of the Sixth European Conference on Diamond, Diamond-like and Related Materials* (Barcelona, 1995, p. 8094).

Translated by R. M. Durham

Comparative emission characteristics of a negative hydrogen ion source with a reflex discharge with and without Cs

V. P. Goretskiĭ, A. V. Ryabtsev, I. A. Soloshenko, A. F. Tarasenko, and A. I. Shchedrin

Institute of Physics, Ukrainian National Academy of Sciences, 252650 Kiev, Ukraine
(Submitted December 19, 1997)
Zh. Tekh. Fiz. **69**, 102–109 (April 1999)

A theoretical and experimental investigation is made to determine how cesium in the volume and at the surfaces of an ion source influences its emission characteristics. It is shown that under the real conditions of an ion source, cesium in the volume makes an appreciable influence to the kinetic processes but barely affects the H^- ion current extracted from the source. However, cesium at the surface of the source increases the H^- ion current severalfold even when the $H–H^-$ conversion efficiency is low ($\gamma \approx 10^{-3}$). The theoretical conclusions show good agreement with the experimental data obtained in the present study. © *1999 American Institute of Physics.* [S1063-7842(99)01804-8]

INTRODUCTION

The present paper is a continuation of a series of studies[1–4] on the kinetic processes in negative hydrogen ion sources with a reflex discharge. Here we report theoretical and experimental investigations of the influence of cesium in the volume and at the surface of an ion source on its emission characteristics. It has been shown in various experimental studies (see, for example, Refs. 5–7) that the addition of cesium vapor to plasma sources substantially increases the H^- ion current density and improves the gas economy of the source. In order to make optimum use of this effect, we need to identify its mechanism. It is probably attributable to the enhanced efficiency of the H^- ion formation processes in the discharge volume or to the additional formation of H^- ions at the surface of the anode which has a low work function because of the cesium film. The influence of cesium in volume processes has been examined in various theoretical studies. Baksht *et al.*[8–11] investigated a cesium–hydrogen discharge at high hydrogen pressure and showed that a comparatively small amount of added cesium ($N_{Cs}/N_{H_2}=10^{-3}$ -10^{-2}) significantly alters the plasma parameters in the gas mixture. The concentration of negative hydrogen ions may reach $N_{H^-}=10^{13}\,cm^{-3}$ (Ref. 9). The generation of negative ions at the cesiated surfaces of the discharge chamber and aspects of H^- ion extraction were not considered in Refs. 8–11.

In contrast to Refs. 8–11, Fukumasa and Niitani[12] examined only the influence of cesium via surface effects involving the conversion of H, H^+, and H_2^+ to H^- at the cesiated anode of a two-chamber source. These authors note that surface conversion may increase the concentration of H^- ions severalfold compared with a pure hydrogen discharge.

Here we make numerical calculations allowing for the influence of cesium on the volume mechanisms for the formation of H^- ions and for the formation of H^- ions at the cesiated surface of the anode. We show that under the real conditions of an ion source, cesium in the volume makes a considerable contribution to the kinetic processes but barely influences the H^- ion current extracted from the source. However, even when the $H–H^-$ conversion efficiency is low ($\gamma \approx 10^{-3}$), cesium at the surface of the source increases the H^- ion current severalfold. These theoretical conclusions show good agreement with the experimental data obtained.

EXPERIMENTAL INVESTIGATIONS OF SOURCE EMISSION CHARACTERISTICS

The source is shown schematically in Fig. 1. The electron source was a heated tungsten cathode *1*, 2 mm in diameter. A diaphragm *2* limiting the radial size of the discharge and the plasma column *6* was positioned in front of the anode chamber *3*. The electrons which ionize the gas were reflected by an anticathode *7* at the same potential as the cathode. Gas was admitted into the anode chamber via equidistant apertures *5*. A maximum field of up to 2 kOe was directed along the axis of the system. The ions were extracted through a slit by the field of the electrode *4*. The diameter of the discharge chamber was 2.5 mm and the anode diameter was 5 mm. This ratio of parameters gave the maximum negative ion yield. The hydrogen pressure in the source chamber was varied between 2×10^{-2} and 2×10^{-1} Torr, the discharge voltage was varied between 100 and 200 V, and the discharge current between 1 and 10 A. The extracting voltage was varied between 8 and 14 kV. The plasma parameters were measured in order to compare the experimental results with the calculations. The plasma density in the column obtained from measurements of the ion current to the anticathode increased proportionately with the discharge current and reached $10^{14}\,cm^{-3}$ (the electron temperature in the discharge column varied between 2 and 6 eV when the pressure varied between 2×10^{-1} and 2×10^{-2} Torr). The plasma parameters outside the column were measured using a thin cylindrical probe inserted in the plasma through the emission slit and oriented perpendicular to the magnetic field. The plasma density at the edge was approximately an order of magnitude lower than that in the column, and the electron temperature at $p\approx1\times10^{-1}$ Torr

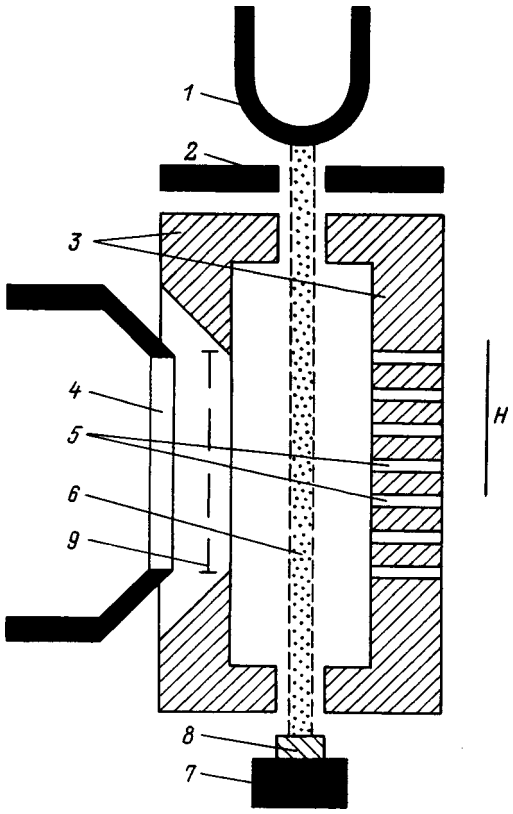

FIG. 1. Schematic of ion source.

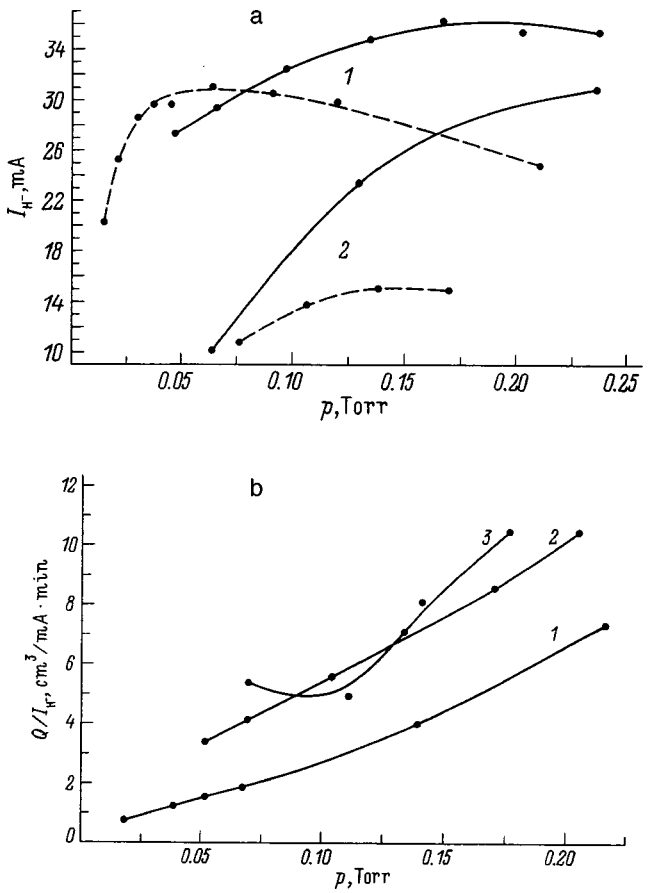

FIG. 2. Emission characteristics of source: a — extracted H^- ion current versus pressure in source chamber (*1* — with cesium, *2* — without cesium, solid curve — 1.5×40 mm emission slit, dashed curve — 0.7×40 mm emission slit); b — gas consumption per unit extracted H^- ion current versus pressure (*1, 2* — source with cesium for emission slits of 0.7×40 or 1.5×40 mm, respectively; *3* — source without cesium , 0.7×40 mm emission slit).

was around 1 eV. Even these data show that this discharge is an analog of the discharge in the well-known two-chamber systems developed for neutral particle injectors. Inside the discharge the fast electrons create a dense plasma and produce vibrationally excited H_2 molecules, and outside the column optimum conditions exist for the formation of H^- ions. The electrons are cooled by diffusion perpendicular to the magnetic field.

Cesium was introduced by the sputtering of a cesium dichromate pellet *8* placed on the reflector *7* (Fig. 1). When the discharge was initiated in the presence of this pellet, the discharge voltage dropped to 50 V and the negative ion current extracted from the source decreased. After several hours, the discharge voltage returned to the levels typical of pure hydrogen and the H^- ion current increased. Measurements were also made under these conditions.

Figure 2a gives the H^- ion current extracted from the source as a function of pressure at an extraction voltage of 14 kV. It can be seen that the addition of cesium appreciably alters the emission characteristics of the source. The main changes are as follows: 1) the maximum density of the H^{-1} ion current increases; 2) the gas economy is improved, especially at low pressure (Fig. 2b); 3) the admission of cesium has less effect as the slit width increases.

Attention should be drawn to a change in the behavior of these curves: whereas in pure hydrogen there is an optimum pressure at which the best gas economy is achieved ($p \approx 10^{-1}$ Torr), when cesium is added it is inversely proportional to the H_2 pressure. The best gas economy is ≈2% which is an order of magnitude higher than that in pure hydrogen.

Even these experimental results indicate that cesium has a predominantly surface effect. This is evidenced mainly from the fact that an increase in the slit width with added cesium does not lead to a proportional increase in the extracted H^- ion current (Fig. 2, curves *1*). Without cesium however, the extracted current is always proportional to the slit width (curves *2*). Another two factors also indicate the surface nature of the effect: 1) cesium has the greatest effect at low gas pressures, when volume processes are less important; 2) with added cesium, the extracted H^- ion current depends more weakly on pressure compared with a pure hydrogen discharge.

However, for confirmation we carried out special experiments in which the emission slit was shut off by a metal grid *9* isolated from the anode (Fig. 1). The metal grid had a spacing of 0.2 mm and a transmission coefficient of 0.8. The voltage at this grid relative to the anode was varied between −25 and +25 V. Figure 3 gives the H^- ion current as a function of the grid voltage at various pressures for pure hydrogen and with added cesium. Without cesium the maximum negative ion current is observed at grid voltages of the

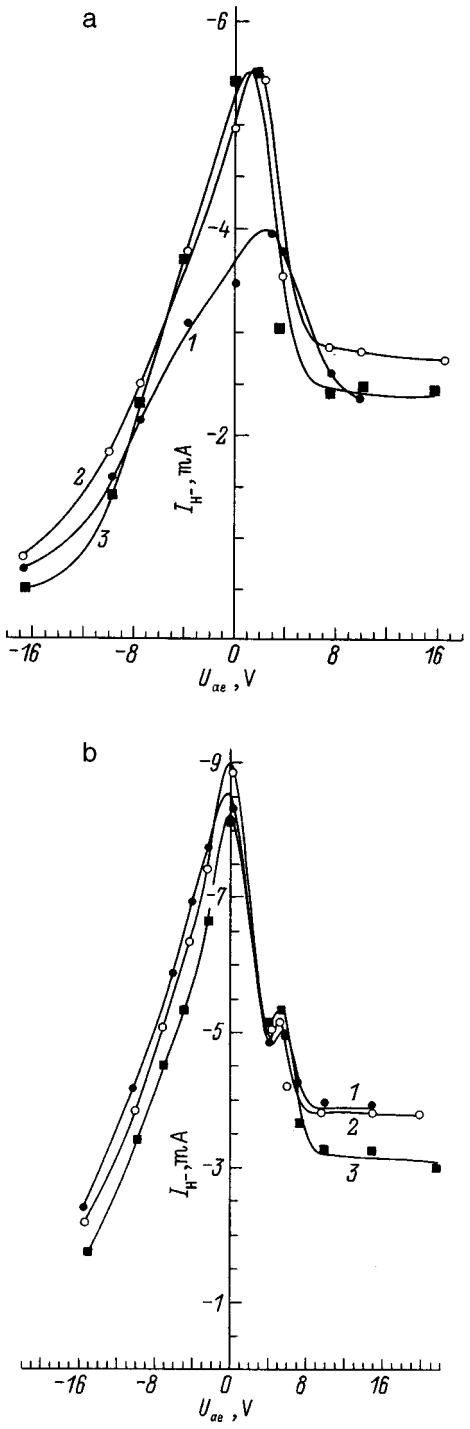

FIG. 3. Negative hydrogen ion current versus potential of emission electrodes (0.7×40 mm emission slit covered by grid): a — without cesium, b — with cesium, source pressure: *1* — 3.74×10^{-2}, *2* — 5.6×10^{-2}, and *3* — 8.7×10^{-2} Torr.

order of 4–5 V. In the discharge with added cesium two maxima are observed: a clearly defined peak at zero potential and a less defined peak at around 5 V. However, the maximum electron current extracted from the emission slit corresponds to an electrode potential of the order of 4–5 V in both cases. These results may be interpreted as follows. Probe measurements show that the plasma potential relative to the anode is positive in the discharge region and varies between 3 and 5 V, depending on the pressure, i.e., the electric field between the column and the anode is a retarding field for the electrons and H^- ions. At a grid voltage of the order of the plasma potential the electric field in the plasma is partially compensated, and the conditions for the extraction of electrons and H^- ions formed as a result of volume processes become more optimal than those at zero potential. If the grid potential is increased further, the electrons and negative ions begin to be trapped by the grid. This can explain the existence of a maximum negative ion current a voltages of 4–5 V without cesium and the second peak in the presence of cesium. The maximum at zero grid potential in a cesium–hydrogen discharge indicates that most of the H^- ions are created at the cesiated surface of the emission slit at zero potential.

NUMERICAL SIMULATION

In order to determine the current density of the negative H^- ions and calculate the concentrations of charged (n_e, N_{H^-}, $N_{H_2^+}$, N_{H^+}, N_{Cs^+}, $N_{Cs_2^+}$), and neutral (N_{H_2}, N_H, $N_{H_2(v)}$, N_{Cs}) mixture components, we solved a system of transport equations (1)–(3) together with the Boltzmann equation[13,14]

$$\frac{\partial N_i}{\partial t}+\operatorname{div}\Gamma_i=\sum_i k_i N_i+\sum_{i,j} k_{ij}N_iN_j, \tag{1}$$

$$\Gamma_i=\mu_i E N_i-D_i\frac{\partial N_i}{\partial r} \tag{2}$$

(μ_i has the appropriate charge sign),

$$n_e+N_{H^-}\approx N_{H_2^+}+N_{H^+}+N_{Cr^+}+N_{Cs_2^+}, \tag{3}$$

where E is the radial electric field, and μ_i and D_i are the mobilities and diffusion coefficients of the charged components, respectively.

We assumed that the plasma was uniformly distributed along the axis of the discharge chamber and in the azimuthal direction.

The spatial distribution of the neutral components was assumed to be uniform along all the axes since their mean free paths are of the order of or greater than the chamber radius under the experimental conditions. Thus, for these components we find $\operatorname{div}\Gamma_i=0$.

The rates of the inelastic processes

$$k_{ij}=\sqrt{\frac{2q}{m}}\int_0^\infty \varepsilon Q_{ij}f_0(\varepsilon)d\varepsilon \tag{4}$$

were calculated using the electron energy distribution function $f_0(\varepsilon)$. In Eq. (4), ε is the electron energy (in electron volts), m is the electron mass (in grams), $q=1.602\times10^{-12}$ erg/eV, and $Q_{ij}(\varepsilon)$ is the cross section of the corresponding elastic process (in centimeters squared).

The electron distribution function in the discharge was determined from the Boltzmann equation assuming that the electron density in the column depends weakly on the radius. It was shown in our earlier studies[1–4] that in pure hydrogen the electron distribution function in the column has a flat

TABLE I.

Process No.	Process
1	$H_2(\nu)+e\rightarrow H_2^+ +e+e,\ \nu=0,\ldots,14$
2	$H+e\rightarrow H^+ +e+e$
3	$H_2+e\rightarrow H+H+e$
4	$H+H(\text{wall})\rightarrow H_2$
5	$H_2^+ +e\rightarrow H+H$
6	$H_2+e\leftrightarrow H_2(\nu)+e,\ \nu=1,2,3$
7	$H_2+e\rightarrow H_2^*(B^1\Sigma_u^+, c^1\Pi_u)+e\rightarrow H_2(\nu)+e+\hbar\omega,$ $\nu=1,\ldots,14$
8	$H_2(\nu)+H\rightarrow H_2(\nu')+H$
9	$H_2(\nu)+\text{wall}\rightarrow H_2(\nu'),\ \nu=1,\ldots,14,\ \nu'=0,\ldots,\nu$
10	$H_2(\nu)+e\rightarrow H^- +H,\ \nu=1,\ldots,14$
11	$H+e\rightarrow H^-$
12	$H_2^+ +e\rightarrow H^- +H^+$
13	$H^- +H_2^+ \rightarrow H+H_2$
14	$H^- +H^+ \rightarrow H+H$
15	$H^- +H_2\rightarrow H+H_2+e$
16	$H^- +H\rightarrow H+H+e$
17	$H^- +e\rightarrow H+e+e$
18	$H_2^+ +e(\text{wall})\rightarrow H_2$
19	$H^+ +e(\text{wall})\rightarrow H$
20	$Cs+e\rightarrow Cs^+ +e$
21	$Cs^+ +Cs+M\rightarrow Cs_2^+ +M$
22	$Cs_2^+ +e\rightarrow Cs+Cs$
23	$Cs^+ +H^-\rightarrow Cs+H$
24	$Cs_2^+ +H^-\rightarrow Cs+Cs+H$
25	$Cs^+ +e(\text{wall})\rightarrow Cs$
26	$H+e(\text{wall})\rightarrow H^-$

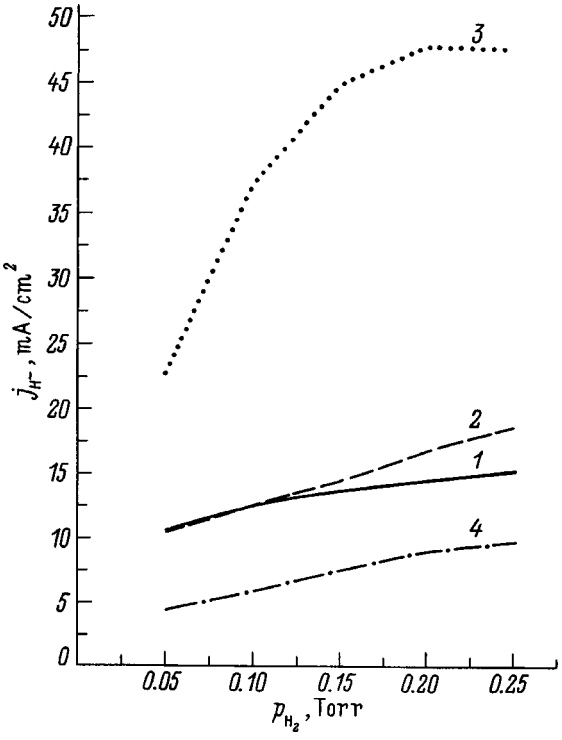

FIG. 4. Density of H^- ion current at anode surface versus pressure in source: *1* — pure hydrogen discharge, *2* — discharge with added cesium allowing only for volume reactions, *3* — discharge with cesium only at the surface, *4* — the same as *2*; U_p, V: *1–3* — 120, *4* — 50.

plateau between the beam input energy of ~ 100 and ~ 20 eV and at low energies is almost Maxwellian (T = 3–5 eV). Between the column and the anode the electron distribution function has no high-energy component because of the strong magnetization of the electrons along the radius and the electron temperature is lower than that in the column ($T_e \leqslant 1$ eV). As a result, the electron distribution function outside the discharge chamber was assumed to be Maxwellian f_{0m} with a radially dependent temperature obtained from the energy balance equation

$$\frac{3}{2}V_e\frac{dT_e}{dr}=-\varepsilon_\nu\sqrt{\frac{2q}{m}}N_{H_2}\int_0^\infty \varepsilon f_{0m}(\varepsilon)Q_\nu(\varepsilon)d\varepsilon,$$

$$f_{0m}=\frac{2}{\sqrt{\pi}}T_e^{-3/2}e^{-\frac{\varepsilon}{T_e}}. \qquad (5)$$

Here V_e is the electron drift velocity at right angles to the magnetic field with allowance for the ambipolar field, $Q_\nu(\varepsilon)$ is the cross section for excitation of the first vibrational level of H_2 (since vibrational excitation is the main channel for loss of electron energy between the column and the anode), and ε_ν is the vibrational quantum energy (in electron volts).

The mobilities and diffusion coefficients were calculated with allowance for the magnetization. We assumed that the ions have a Maxwellian distribution throughout the discharge chamber. The main mechanism determining the transport coefficients for the electrons at pressure $p\sim 0.1$ Torr and plasma densities of $\sim 10^{13}$–10^{14} cm^{-3} is scattering by ions. As in Ref. 2, we assumed that the temperature of atomic hydrogen T_H and the negative ions T_{H^-} is 4×10^3 K. The temperature of the other ions and atomic components in the mixture is close to room temperature.

Equations (1)–(3) and the Boltzmann equation were solved taking into account the elementary processes listed in Table I. The transport coefficients k_1 describe linear processes in terms of concentration (Nos. 4, 9, and 18–20 in Table I).

It should be noted that the calculations neglected the mechanism for the formation of negative ions via Rydberg states, since numerical estimates show that under the present experimental conditions the contribution of this mechanism does not exceed a few percent. The formation of negative H^- ions in the volume by dissociative attachment of electrons to CsH molecules was also neglected since the CsH concentration is low under our discharge conditions. The most optimistic estimates show that the rate of H^- production by this mechanism is two orders of magnitude lower than that via process No. 10 (see Table I).

CALCULATED RESULTS AND DISCUSSION OF THE EXPERIMENT

As was noted in the Introduction, the main purpose of this study was to determine how cesium in the volume of the discharge chamber and adsorbed on the surface influences the emission characteristics of a negative hydrogen ion source with a reflex discharge. As a result we made a numerical simulation of a reflex discharge for three main scenarios: 1) a pure hydrogen discharge (taking into account processes 1–19 in Table I; 2) cesium is present only in the

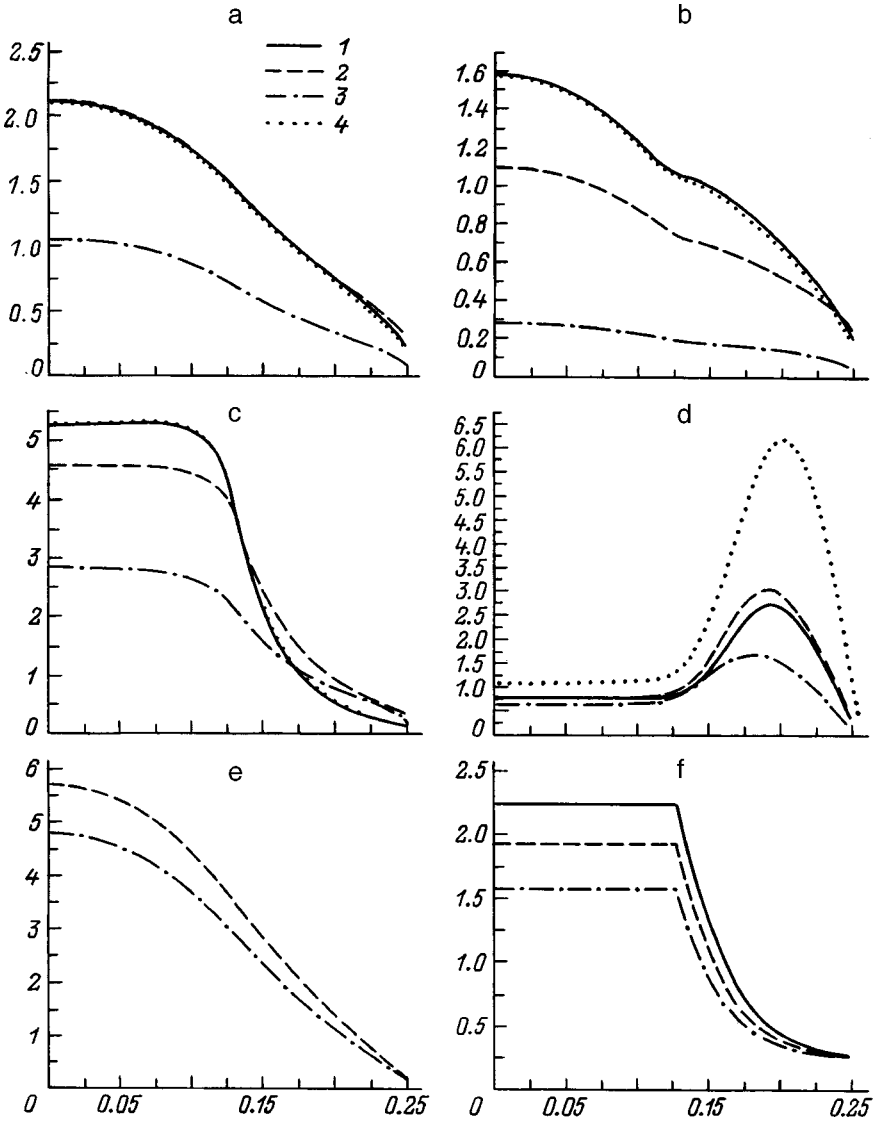

FIG. 5. Distribution of the density of various plasma components and temperature along the radius of the discharge column: a — electron density ($10^4\,cm^{-3}$); b — H^+ ion density ($10^{14}\,cm^{-3}$); c — H_2^- ion density ($10^{13}\,cm^{-3}$); d — H^- ion density ($10^{11}\,cm^{-3}$); e — density of Cs^+ ions ($10^{13}\,cm^{-3}$), electron temperature (eV). Curves *1–4* — the same as in Fig. 4.

volume of the discharge chamber (processes 1–25); 3) cesium is present only at the surface (processes 1–19, and 26). Conversion of H^+ to H^- ions at the cesiated surface was neglected, since the concentration of atomic hydrogen was much higher than the H^+ concentration under these discharge conditions.

Figure 4 gives the current density of negative hydrogen ions j_{H^-} at the anode, plotted as a function of pressure for the three cases listed above. It can be seen from Fig. 4 (curve *2*) that the addition of cesium vapor to the volume of the discharge chamber under conditions where the discharge voltage is artificially maintained merely leads to a negligible increase in the current j_{H^-} at a higher pressure compared with pure hydrogen (curve *1*). The main mechanism responsible for the formation of negative hydrogen ions in this case, as in a pure hydrogen discharge, is the dissociative attachment of electrons to vibrationally excited hydrogen molecules (process No. 10). A slight increase in the negative ion current when Cs is added is attributable to the following factors. In volume processes cesium plays both a positive and a negative role. Since cesium has a low electron excitation and ionization energy, the number of electrons having energies higher than the Cs excitation energy decreases. The electron temperature T_e ($T_e=-1/(d\ln(f_0/d\varepsilon))$, which is meaningful up to $\leqslant 20$ eV) then decreases (Fig. 5f), in agreement with the experimental data. However, a drop in the electron temperature and in the number of fast electrons slows the rate of formation of vibrationally excited H_2 molecules and therefore H^- ions. At the same time, a drop in T_e leads to a reduction in the rate of electron detachment (process No. 17, the cross section for this process depends strongly on temperature) and also reduces the rate of H_2 dissociation and thus the concentration of atomic hydrogen (Fig. 6), which is actively involved in removing the H_2 vibrational excitation. These last two factors promote an increase in the concentration of H^- ions. These negative and positive effects of introducing Cs into the volume almost cancel out when the same discharge voltage is artificially maintained (Fig. 4). The actual lowering of the voltage to 50 V observed in experiments where cesium is added to the volume leads to a reduction in the H^- ion current (Fig. 4, curve *4*). Thus, in agreement with the experiments the numerical simulation indicates that cesium in the volume can-

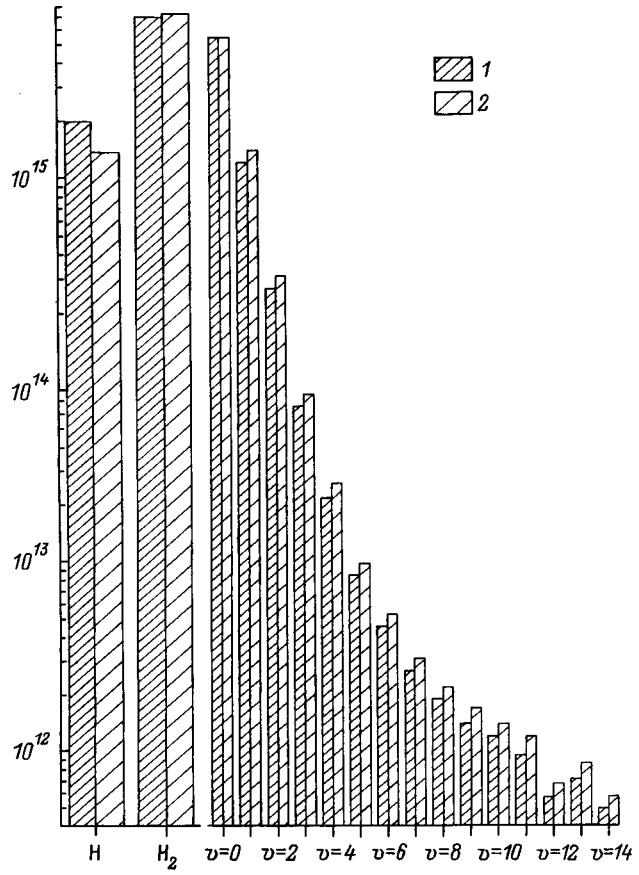

FIG. 6. Comparative concentrations of atomic and molecular hydrogen and also various vibrationally excited levels of the H_2 molecule: *1* — pure hydrogen discharge, *2* — discharge with cesium only in the volume.

not increase the current density of the H^- ions extracted from the source.

In contrast, allowance for conversion of H to H^- at the anode with adsorbed cesium even with a conversion efficiency $\gamma = 10^{-3}$ leads to an increase in the H^- current (Fig. 4, curve *3*) and the H^{-1} concentration in the volume (Fig. 5e) by a factor of 2–4. Without making a detailed comparison between the calculations and the experiment, we note that this increase agrees with the experimentally observed change in the H^- current (Fig. 2). We also note that under the experimental conditions no special measures were taken to optimize the cesium coating of the anode surface. However, it follows from Ref. 15 that the conversion efficiency for ion sources usually varies between 10^{-3} and 10^{-1} depending on the temperature of the hydrogen atoms and the cesium coverage of the surface.

The concentration of negative ions in pure hydrogen and with added cesium reaches a maximum in the region between the discharge column and the anode. This important factor for the experiments can be explained as follows. The value of N_{H^-} is determined by the H^- production and loss processes. The main mechanism responsible for the loss of H^- is detachment of electrons as a result of collisions between H^- and fast plasma electrons (process No. 17). The cross section for this process depends on energy, as has already been noted, and exhibits threshold behavior: the cross section is zero at electron energies $\varepsilon \leqslant 1.25$ eV and at energies of ~ 10 eV reaches $\sim 2 \times 10^{-15}$ cm^2. Since the electron temperature outside the column decreases along the radius (Fig. 5f), the rate of this process drops sharply, which leads to an increase in the H^- concentration despite some decrease in their rate of formation. If it is artificially assumed that the rate of process No. 17 is constant, the maximum disappears. It can be seen from Fig. 5e that for a discharge column of radius $R_k = 0.125$ cm the maximum H^- ion concentration is achieved for $R/R_k \approx 1.4$. The experimental results given for various diameters of the anode chamber R_a show good agreement with the calculations: the maximum negative ion current extracted from the source without cesium is obtained for $R_a/R_k = 2$. It was noted in Ref. 4 that some difference between the numerical values of the calculated and experimental negative ion current extracted from the source can be explained first by the fact that the area of the plasma emission surface near the slit is considerably greater than the surface area of the slit; second, for these pressures the diffusion approximation used in the calculations to analyze the H^- ion motion is at the limit of its validity.

A reduction in the H^- ion current observed experimentally at high pressures may be caused by charge transfer of negative ions with the gas jet leaving the emission slit of the source. Estimates show that at high pressure this process may result in the loss of up to 30% of H^- ions leaving the slit, which is quite sufficient for the appearance of a maximum on the experimentally measured dependence $j_{H^-}(p)$.

It can be concluded from these experimental results and calculations that cesium in the volume of the ion source cannot increase the H^- ion current. The observed increase in the H^- ion current is caused by conversion of hydrogen atoms at the cesiated surface of the anode. In other words, the addition of cesium converts a volume-type of H^- ion source into a surface plasma type.

This work was supported by the Fund for Fundamental Research of the Ministry of Science of the Ukraine (Agreement No. F4/342-97-32, 1997).

[1] P. M. Golovinskiĭ, V. P. Goretskiĭ, A. V. Ryabtsev *et al.*, Zh. Tekh. Fiz. **61**(10), 46 (1991) [Sov. Phys. Tech. Phys. **36**, 1102 (1991)].
[2] P. M. Golovinskiĭ, V. P. Goretskiĭ, I. A. Soloshenko *et al.*, Zh. Tekh. Fiz. **63**(9), 46 (1993) [Tech. Phys. **38**, 762 (1993)].
[3] V. P. Goretskiĭ, A. V. Ryabtsev, I. A. Soloshenko *et al.*, Zh. Tekh. Fiz. **64**(7), 152 (1994) [Tech. Phys. **39**, 715 (1994)].
[4] V. P. Goretskiĭ, A. V. Ryabtsev, I. A. Soloshenko *et al.*, Zh. Tekh. Fiz. **66**(2), 88 (1996) [Tech. Phys. **41**, 161 (1996)].
[5] Y. Okumara, M. Hanada, T. Inoue *et al.*, in *Proceedings of the Fifth International Symposium on Production and Neutralization of Negative Ions and Beams*, Brookhaven, New York (1990), p. 169.
[6] S. R. Walter, K. N. Leung, and W. B. Hunkel, J. Appl. Phys. **64**, 3424 (1988).
[7] Yu. I. Bel'chenko, G. I. Dimov, and V. G. Dudnikov, Preprint No. 39-73 [in Russian], Institute of Nuclear Physics, Siberian Branch of the Academy of Sciences of the USSR, Novosibirsk (1973).
[8] F. G. Baksht and V. G. Ivanov, Pis'ma Zh. Tekh. Fiz. **12**, 672 (1986) [Sov. Phys. Tech. Phys. **12**, 278 (1986)].
[9] F. G. Baksht, L. I. Elizarov, and V. G. Ivanov, Fiz. Plazmy **16**, 854 (1990) [Sov. J. Plasma Phys. **16**, 497 (1990)].
[10] F. G. Baksht and V. G. Ivanov, Zh. Tekh. Fiz. **62**(2), 195 (1992) [Sov. Phys. Tech. Phys. **37**, 223 (1992)].
[11] F. G. Baksht and V. G. Ivanov, Pis'ma Zh. Tekh. Fiz. **23**(1), 26 (1997) [Tech. Phys. Lett. **23**(1), 20 (1997)].

[12] O. Fukumasa and E. Niitani, in *Proceedings of the Seventh International Symposium on Production and Neutralization of Negative Ions and Beams*, Woodbury, New York (1995), p. 187.
[13] I. P. Shkarofsky, T. W. Johnston, and M. P. Bachynskii, *The Particle Kinetics of Plasmas* [Addison–Wesley, Reading, Mass., 1966; Atomizdat, Moscow, 1969].
[14] V. P. Goretskiĭ, A. V. Ryabtsev, I. A Soloshenko *et al.*, Fiz. Plazmy **20**, 836 (1994) [Plasma Phys. Rep. **20**, 755 (1994)].
[15] M. Seidl, H. L. Cui, J. D. Isenbery *et al.*, in *Proceedings of the Sixth International Symposium on Production and Neutralization of Negative Ions and Beams*, Brookhaven, New York (1992), p. 25.

Translated by R. M. Durham

BRIEF COMMUNICATIONS

Influence of impurities on the photoluminescence of modified InP crystals

F. B. Baimbetov and N. G. Dzhumamukhambetov

Atyrau University, Atyrau, Kazakhstan
(Submitted March 18, 1998)
Zh. Tekh. Fiz. **69**, 110–111 (April 1999)

It is shown that a new emission band at 1.35 eV (77 K) is observed in the photoluminescence spectra of laser-modified InP crystals, irrespective of the type of impurity or its concentration in the initial crystal. It is established that the emergence of this band is a common property of the lattice defect state of the modified semiconductors. © *1999 American Institute of Physics.* [S1063-7842(99)01904-2]

The luminescence properties of modified InP crystals has been reported previously.[1] It has been shown that a new band with a maximum at an energy of 1.35 eV is observed in the photoluminescence spectra at 77 K on account of the random distribution of structural defects after laser treatment. The objective of the present study is to investigate the influence of the dopant impurity in the initial crystal on the luminescence properties of the laser-modified crystals.

The original materials used to prepare the modified crystals were *n*-type and *p*-type InP: nominally undoped, $n=(1-2)\times10^{16}\,\mathrm{cm}^{-3}$, doped with tin, $n=2\times10^{18}\,\mathrm{cm}^{-3}$, and doped with zinc, $p=(7-8)\times10^{16}\,\mathrm{cm}^{-3}$. The laser treatment procedure applied to the crystals is described in Ref. 1. Prior to the measurements the crystals were etched in a freshly prepared H_2O:HNO_3:HCl (6:3:1) polishing etchant at room temperature.

The luminescence spectra of the initial crystals at 77 K exhibited typical emission bands with energy maxima of 1.415 eV, 1.405 eV, 1.371 eV, and 1.113 eV. At room temperature the luminescence spectra of these crystals consists of only one broad band with an energy maximum of 1.344–1.347 eV. The nature of these bands is discussed in Refs. 2 and 3.

After irradiation by trains of pulses with a total energy density of $100\,\mathrm{J/cm}^2$ the polished surfaces of the crystals undergo a visible change, irrespective of which dopant is used. At 77 K the photoluminescence spectra acquire a new emission band with an energy maximum of 1.350 eV. However, its intensity is much lower than the intensity of the edge band of the initial crystals. As the total irradiation energy density increases, the intensity of this band increases, attaining its maximum value at $E=250-300\,\mathrm{J/cm}^2$. At 77 K its intensity is 5–10 times (for different crystals) the intensity of the edge band of the initial crystal.

The temperature curves of the main parameters of the 1.35 eV band coincide qualitatively for all the crystals, but certain differences are observed quantitatively. For example, the total intensity decreases exponentially as the temperature increases (Fig. 1):

$$\Phi=\Phi_0\exp(-kT/\varepsilon^*),$$

where the characteristic energy ε^* lies in the interval 3.4–4.3 meV for different crystals.

The half-width of this band varies nonmonotonically as the temperature increases (Fig. 2). In the low-temperature range $T<230$ K it increases linearly, attaining its maximum value at $T=230-250$ K (for different crystals). It decreases in the high-temperature range $T>250$ K.

The maximum energy $h\nu_m$ also varies nonmonotonically with increasing temperature (Fig. 3), its curves exhibiting qualitatively the same form for all crystals. However, the energy maximum for modified crystals obtained from as-grown *p*-InP exceeds the corresponding values of the crystals

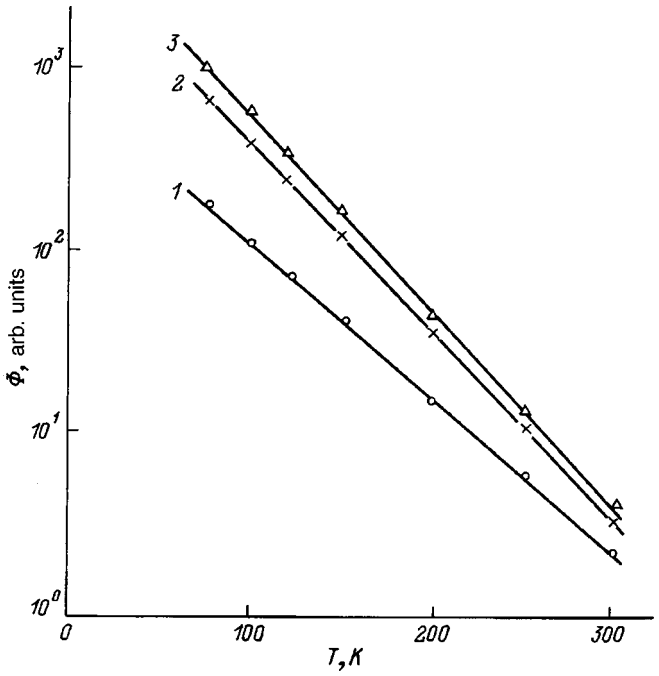

FIG. 1. Temperature dependence of the total intensity of the 1.35 eV band of modified crystals obtained from different grades of initial InP: (*1*) InP⟨Zn⟩; (*2*) InP⟨Sn⟩; (*3*) InP.

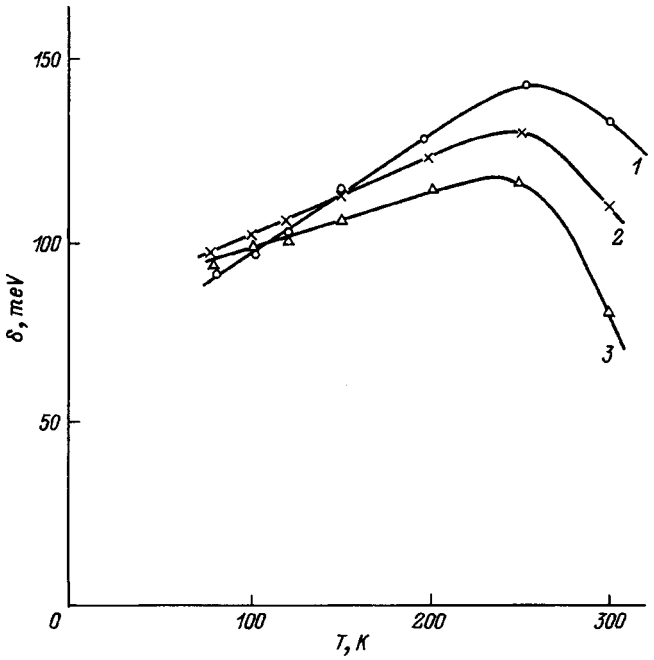

FIG. 2. Temperature dependence of the half-width of the 1.35 eV band of modified crystals obtained from different grades of initial InP: (*1*–*3*) the same as in Fig. 1.

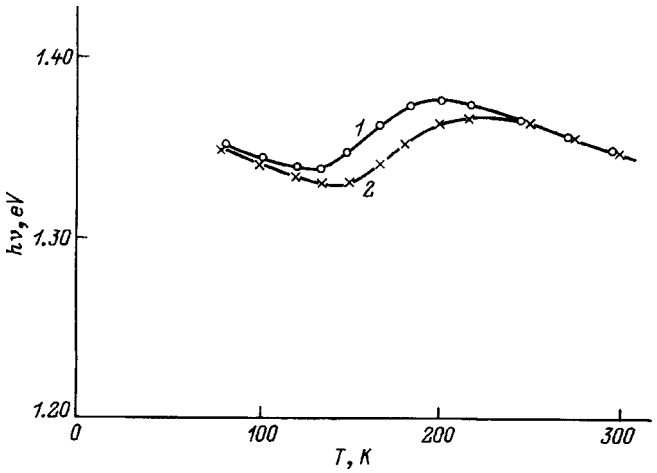

FIG. 3. Temperature dependence of the energy maximum of the 1.35 eV band of modified crystals obtained from initial InP with different types of conductivity: (*1*) *p*-type; (*2*) *n*-type.

obtained from *n*-InP at the same temperature. Here the temperature corresponding to the maximum value $h\nu_m$ for modified crystals prepared from *p*-InP, ~ 120 K, is lower than for crystals obtained from *n*-InP, ~ 140 K.

Apart from their quantitative differences, the temperature curves of the main parameters of the 1.35 eV emission band coincide, irrespective of the dopant used. The nonmonotonic variation of the energy maximum and the half-width as well as the exponential decay of the intensity with increasing temperature indicates that the density of states has tails near the allowed bands, formed there as a result of the random distribution of structural defects of the InP crystal itself after laser treatment.

Consequently, the emergence of a new 1.35 eV band in the photoluminescence spectra of the modified InP crystals at 77 K does not depend on the type of impurity or its concentration in the initial material and is a common property of the defect state of the crystal lattice.

[1] N. G. Dzhumamukhambetov and A. G. Dmitriev, Fiz. Tekh. Poluprovodn. **27**, 641 (1993) [Semiconductors **27**, 356 (1993)].
[2] E. W. Williams and W. Elder, J. Electrochem. Soc. No. 12, 120 (1973).
[3] L. I. Kolesnik, A. M. Loshinskiĭ, A. Ya. Nashel'skiĭ, and S. V. Yakobson, Izv. Akad. Nauk SSSR, Neorg. Mater. **17**(12), 14 (1981).

Translated by James S. Wood

Modeling of the low-temperature production of gas-sensitive tin oxide films

V. V. Kisin, S. A. Voroshilov, V. V. Sysoev, and V. V. Simakov

Saratov State Technical University, 410054 Saratov, Russia;
N. G. Chernyshevski Saratov State University, 410601 Saratov, Russia
(Submitted February 24, 1998)
Zh. Tekh. Fiz. **69**, 112–113 (April 1999)

The deposition of tin oxide films by reactive sputtering of a stoichiometric target is investigated. Conditions are determined for the formation of a crystal layer structure and the acquisition of gas sensitivity without subsequent high-temperature annealing. © *1999 American Institute of Physics.* [S1063-7842(99)02004-8]

Stannic oxide (SnO_2) films prepared in vacuum from a stoichiometric source become sensitive to gases after they are annealed in oxygen.[1] They acquire this property because during deposition some of the oxygen is unavoidably pumped out of the region where the film is formed. As a result, an amorphous or heterophase layer containing Sn, SnO, and SnO_2 is deposited, and its homogenization requires many hours of subsequent treatment at high temperatures,[2,3] making it difficult to integrate the active layer into complex sensing devices such as, for example, smart sensors.

At the same time, homogeneous polycrystalline SnO_2 films can be obtained as the result of a low-temperature process involving the injection of oxygen into the reactor, without having to resort to high-temperature treatment. However, the conditions underlying the formation of a layer possessing a specified stoichiometry are still not very well defined, as evinced by the attempts of different research groups to use gas mixtures with various oxygen contents.[4–6] Obviously, the composition of the growing layer is not solely dependent on the oxygen flow.

The objective of the present study has been to construct and experimentally test a model of the process of low-temperature formation of a gas-sensitive layer by the reactive sputtering of a stoichiometric target with the capability that the model can be used to relate the composition of the deposited layer to the parameters of the technological process and equipment.

We consider the rf-bias reactive sputtering of a SnO_2 target. We assume that the metal and the oxygen are sputtered at the same constant rate f. The Sn atoms are deposited on the walls and substrate, and the oxygen raises the partial pressure p of the oxygen in the chamber. The pressure p is also influenced by the injection of oxygen from the external source. At a flow rate q_0 the number of oxygen molecules in the chamber increases at the rate $N_a q_0 / V_0$, where N_a is Avogadro's number, and V_0 is the volume of one mole under standard conditions. Evacuation causes the number of oxygen molecules to decrease at the rate $pSN_a / P_0 V_0$, where P_0 is atmospheric pressure, and S is the evacuation rate of the pump. The rate S is customarily assumed to be independent of p in the design of industrial vacuum systems. Finally, according to molecular kinetic theory, oxygen atoms are admitted from the region with pressure p at a flux per unit time onto unit surface of the chamber walls and the substrate $F = 2p/(2\pi mkT)^{1/2}$, where m is the mass of the oxygen molecule, k is Boltzmann's constant, and T is the absolute temperature.

Allowing for the fact that free-electron capture takes place in the chemisorption of oxygen and assuming that, owing to the difference in the densities of free charge carriers, oxygen atoms do not stick to SnO as well as they do to Sn and do not stick at all to SnO_2, we obtain the condition for a constant oxygen pressure to exist in the chamber

$$q = \Psi(\sigma + \alpha + \eta\beta),$$

where α, β, and γ are the fractions of the chamber wall and substrate surfaces coated with Sn, SnO, and SnO_2, respectively, η is the sticking coefficient of oxygen to SnO, $q = 1 + q_0 N_a / V_0 f A_t$ is the input oxygen flux, $\sigma = SN_a p / F_0 V_0 A_s$ is the evacuation efficiency, $\Psi = FA_s / 2fA_t$ is the relative flux of oxygen onto the chamber walls and the substrate, and A_t and A_s are the areas of the target and the deposited film, respectively.

The area occupied by the metal increases as a result of the spraying of Sn onto SnO and decreases as a result of

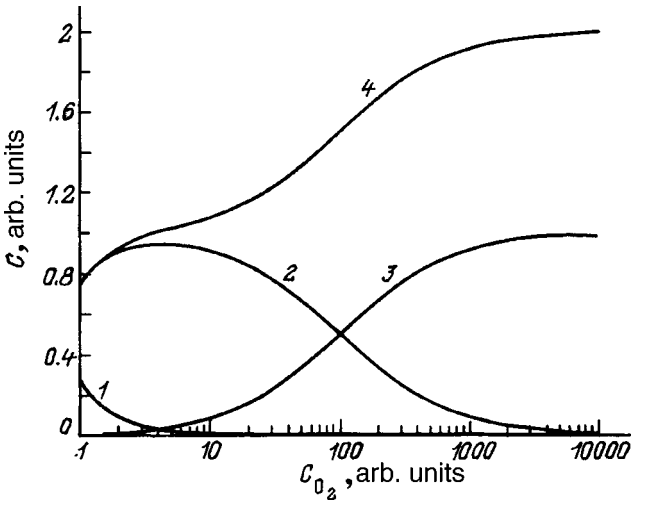

FIG. 1. Dependence of the composition of a tin oxide film on the influx of oxygen for the fractions of: (*1*) Sn; (*2*) SnO, (*3*) SnO_2; (*4*) ratio of O_2 to Sn in the layer.

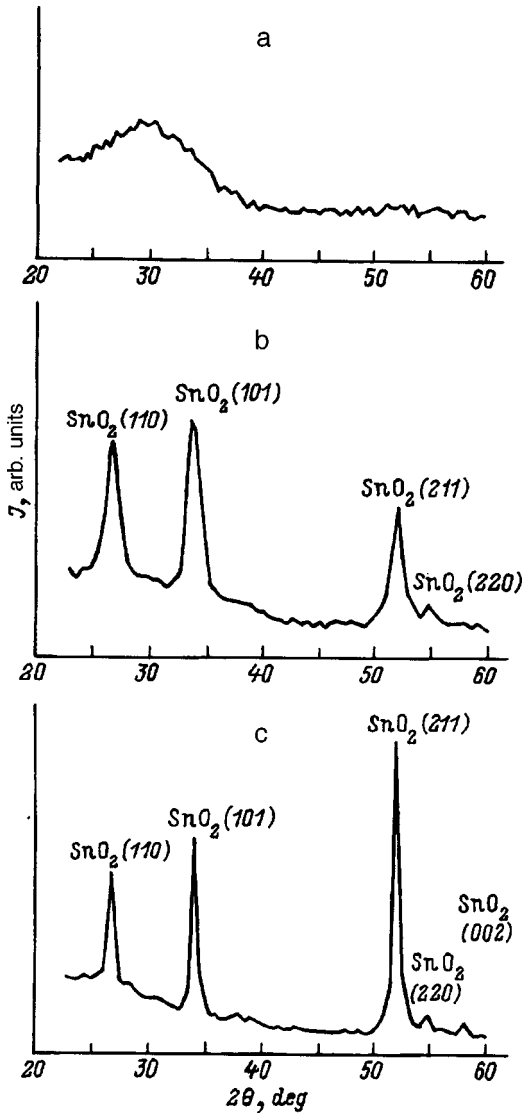

FIG. 2. Diffraction patterns of resulting tin oxide layers: (a) deposition in an Ar atmosphere without heat treatment; (b) deposition with heat treatment in an O_2 flow; (c) deposition in a 1:3 Ar/O_2 gas mixture.

oxidation. The area occupied by SnO_2 increases as a result of the oxidation of SnO and decreases as metal is sprayed onto it. Under steady-state conditions $\Psi\alpha=\beta$ and $\eta\Psi\beta=\gamma$.

The results of calculations of the dependence of the film composition on the oxygen input into the chamber are shown in Fig. 1. Without the external injection of oxygen ($q=1$) the film consists of a mixture of the Sn, SnO, and SnO_2 phases. As the flow rate of oxygen is increased, the resulting film acquires a composition close to SnO, and with a further increase in q its composition is close to SnO_2. If conditions are established on the substrate such as to prevent the crystallization of SnO, the formation of amorphous layers can be expected at low oxygen flow rates, and polycrystalline layers can be expected at high oxygen flow rates.

The model has been tested experimentally by means of an apparatus of the UVP-2M type. Tin oxide films were deposited by magnetron sputtering of a stoichiometric SnO_2 target. During the process the total pressure of the argon–oxygen mixture was 3 mtorr, and the discharge power was approximately 100 W/cm^2. The substrate temperature was set at 170 °C so that the disproportionation of SnO would prevent crystallization of the layer in a lattice differing from a rutile-type lattice. Samples were prepared by deposition in argon with subsequent annealing in an oxygen atmosphere or by deposition in an argon–oxygen mixture. The film thickness was chosen in accordance with the thin-film gas sensitivity model.[7] X-ray structural examinations were performed by means of a DRON-2.0 diffractometer, and the electrical properties were measured in a three-electrode gas sensor configuration.[8]

The films grown in argon were x-ray amorphous (Fig. 2a) and did not exhibit gas sensitivity. A two-hour anneal in an oxygen atmosphere at a temperature of 700 °C induced crystallization of the layers (Fig. 2b), a change in the electrical conductivity, and the suppression of gas sensitivity. The injection of oxygen into the chamber led to the formation of a gas-sensitive polycrystalline film during deposition without additional heat treatment (Fig. 2c). A maximum of the gas sensitivity was observed for films deposited in a 1:3 argon–oxygen mixture, which for our apparatus corresponded to an input flow rate $q=10^3$. A further increase in the flow rate led to an increase in the resistance of the layers as a result of a decrease in the density of intrinsic donors and, hence to degradation of the gas sensitivity.

The proposed model has thus been used successfully to relate the composition of the deposited layer to the parameters of the technological process and equipment, and also to determine the conditions for gas-sensitive layers to be capable of forming at low temperatures.

[1] Y. K. Fang and J. J. Lee, Thin Solid Films **169**, 51 (1989).
[2] G. Williams and G. Coles, Sens. Actuators B **24–25**, 469 (1995).
[3] J.-L. Huang, D.-W. Kuo, and B.-Y. Shew, Surf. Coat. Technol. **79**, 263 (1996).
[4] S. Kaciulis, G. Mattogno, A. Galdikas *et al.*, J. Vac. Sci. Technol. A **14**, 3164 (1996).
[5] M. Di Giulio, A. Serra, A. Tepore *et al.*, Mater. Sci. Forum **203**, 143 (1996).
[6] S. K. Andreev, L. I. Popova, V. K. Gueroguiev, and E. B. Manolov, Vacuum **47**, 1325 (1996).
[7] V. V. Kisin, V. V. Sysoev, V. V. Simakov, and S. A. Voroshilov, in *Proceedings of the Tenth European Conference on Solid State Transducers—''Eurosensors X,''*, Vol. 3 (Leuven, Belgium, 1996), pp. 977–980.
[8] V. V. Kisin, S. A. Voroshilov, V. V. Sysoev, and V. V. Simakov, Pribory Tekh. Éksp., No. 5, pp. 178–181 (1995).

Translated by James S. Wood

Generalization of the Fresnel law to the case of a pressure-wave-induced displacement of the interface between acoustic media

V. A. Pozdeev
(Submitted March 18, 1998)
Zh. Tekh. Fiz. **69**, 114–115 (April 1999)

The problem of interaction between a nonsteady-state pressure wave and a moving interface between acoustic media is analyzed and solved for the first time with allowance for a finite displacement of the interface induced by the wave. An analytic solution is obtained using a nonlinear time transformation method. Expressions are obtained for the law of motion of the interface, and for the reflected and transmitted waves as a function of the time profile of the incident wave and the acoustic characteristics of the media. © *1999 American Institute of Physics.* [S1063-7842(99)02104-2]

The interaction of an acoustic wave with a fixed interface between acoustic media is described by the well-known Fresnel formulas.[1] Isakovich[2] showed that a Doppler effect appears when harmonic waves are incident on an interface moving at a constant velocity. The present author[3] solved the problem of interaction between a nonsteady-state pressure wave and a moving interface for a given arbitrary time law. Here we analyze the reflection and propagation of pressure waves at an interface whose displacement is caused by the incident wave.

As in Ref. 3, we shall assume that a plane pressure wave is incident normal to the interface from a first medium having the acoustic resistance $z_1=\rho_1 c_1$, where ρ_1 is the density of the medium and c_1 is the velocity of sound in the medium. The time profile of the incident wave has the form $p_{10}(t)$. When the wave interacts with media which are initially at rest, it becomes displaced, the wave is reflected from the moving interface, and a wave is generated in the second medium whose acoustic resistance is $z_2=\rho_2 c_2$, $z_2\neq z_1$. To be specific, we direct the x coordinate in the opposite direction to the motion of the incident wave and we position the origin at the initial location of the interface. We measure the time from the instant when the front of the incident wave reaches the contact interface. We assume that the acoustic media are nonviscous and will describe their motion by linear equations.

The following boundary conditions are satisfied at the moving interface between the two media:

$$(p_1-p_2)|_{x=h(t)}=0, \quad p_1=p_{10}+p_{11}, \tag{1}$$

$$(v_1-v_2)|_{x=h(t)}=0, \quad v_1=v_{10}+v_{11}, \tag{2}$$

$$v_2|_{x=h(t)}=\frac{dh}{dt}, \tag{3}$$

where p_1 and p_2 are the pressures and v_1 and v_2 are the particle velocities of the media, and $h(t)$ is the (as yet unknown) law of motion of the interface.

We introduce the velocity potentials of the perturbed motion of the media

$$\Phi_1(x,t)=\Phi_{10}(t-x/c_1)+\phi_{11}(t+x/c_1), \tag{4}$$

$$\phi_2(x,t)=\phi_2(t-x/c_2), \tag{5}$$

where ϕ_i are the velocity potentials of the motion of the ith medium ($i=1,\ 2$), ϕ_{10} is the potential of the incident wave, ϕ_{11} is the potential of the reflected wave, and ϕ_2 is the potential of the transmitted wave.

The pressure and velocity components are related to the velocity potentials by the well-known formulas

$$p_i=-\rho_i\frac{\partial\phi_i}{\partial t}, \quad v_i=\frac{\partial\phi_i}{\partial x}. \tag{6}$$

Following Ref. 3, we shall solve the stated problem (1)–(6) by the nonlinear time transformation method.[4,5] Here we give the expressions for the reflected and transmitted waves in the form

$$p_{11}(x,t)=\frac{z_2-z_1}{z_2+z_1}p_{10}(f_+(t_1^0)), \tag{7}$$

$$p_2(x,t)=\frac{2z_2}{z_2+z_1}p_{10}(f_-(t_2^2)), \tag{8}$$

where

$$f_+(t_1^0)=t_1^0-2h(w_+(t_1^0))/c_1,\ t_1^0=t+x/c_1,$$

$$f_-(t_2^0)=t_2^0+(1-c_2/c_1)h(w_-(t_2^0))/c_2,\ t_2^0=t-x/c_2,$$

$$t=w_+(\xi_1)\to t+h(t)/c_1=\xi_1,$$

$$t=w_-(\xi_2)\to t-h(t)/c_2=\xi_2. \tag{9}$$

We can show that expressions (7)–(9) are equivalent to the corresponding results of Ref. 3. We shall use the boundary condition (3), the coupling relation $v_2=p_2/z_2$, and expression (8) to determine the law of motion of the interface between the media as a result of elastic deformations. We obtain

$$\frac{dh}{dt}=\frac{2}{z_1+z_2}P_{10}(t-h(t)/c_1). \tag{10}$$

If the condition $(h(t)/(c_1t))^2\ll 1$ is satisfied, Eq. (10) may be written in the approximation

$$\frac{dh}{dt}-\frac{2}{z_1+z_2}\frac{dp_{10}(t)}{dt}h=\frac{2}{z_1+z_2}p_{10}(t). \tag{11}$$

Equation (11) can be integrated to yield

$$h=\frac{2}{z_1+z_2}\exp\left(-\frac{p_{10}(t)}{c_1(z_1+z_2)}\right)\times\int_0^t\exp\left(\frac{p_{10}(\tau)}{c_1(z_1+z_2)}\right)p_{10}(\tau)d\tau. \tag{12}$$

We shall analyze an incident wave $p_{10}(t)=p_0\exp(-\alpha t)$, where p_0 is the wave amplitude and α is the time constant. In this case, the initial equation (10) can be integrated exactly:

$$h=-\frac{c_1}{\alpha}\ln(1+\beta_0(e^{-\alpha t}-1)/c_1), \tag{13}$$

where $\beta_0=2p_0/(z_1+z_2)<1$.

If we confine ourselves to the time interval $(\alpha t)^2\ll 1$, Eq. (13) gives

$$h\approx\beta_0 t[1-(\alpha t)(1-\beta_0/c_1)/2+\dots], \tag{14}$$

$$p_{11}\approx\frac{z_2-z_1}{z_2+z_1}p_0\exp(-\alpha_1 t_1^0), \tag{15}$$

$$p_2\approx\frac{2z_2}{z_2+z_1}p_0\exp(-\alpha_2 t_2^0), \tag{16}$$

where

$$\alpha_1\approx\alpha\left(1-\frac{\beta_0}{c_1}\right)\Big/\left(1+\frac{\beta_0}{c_1}\right),$$

$$\alpha_2\approx\alpha\left(1-\frac{\beta_0}{c_1}\right)\Big/\left(1-\frac{\beta_0}{c_2}\right).$$

It can be seen from expressions (14)–(16) that $\alpha_1<\alpha$ is always satisfied and consequently, the reflected wave is stretched compared with the incident wave. When $c_2<c_1$, $\alpha_2>\alpha$, the transmitted wave is stretched relative to the incident one whereas when $c_2>c_1$, $\alpha_2<\alpha$, the transmitted wave is compressed.

It is interesting to consider the inverse problem of determining the form of the incident wave when the law governing motion of the interface between the media is known. Let us assume that

$$h(t)=v_0(t)-a_0t^2/2,\quad 0\leqslant t\leqslant 2v_0/a_0. \tag{17}$$

Then, in accordance with Eq. (10), we obtain

$$p_{10}(t)=\frac{c_1(z_1+z_2)}{2}\left[1-(1-m_1)\sqrt{1+\frac{2a_0t}{c_1(1-m_1)^2}}\right]\approx\frac{v_0(z_1+z_2)}{2}\left[1-\frac{a_0t}{v_0(1-m_1)}+\dots\right],$$

$$m_1=\frac{v_0}{c_1}. \tag{18}$$

Thereafter, expressions for the reflected and transmitted waves can easily be derived using expressions (7) and (8).

[1] E. Skudrzyk, *The Foundations of Acoustics; Basic Mathematics and Basic Acoustics* [Springer-Verlag, New York, 1971; Mir, Moscow, 1976, 250 pp.].
[2] I. A. Isakovich, *General Acoustics* [in Russian], Nauka, Moscow (1973), 496 pp.
[3] V. A. Pozdeev, Pis'ma Zh. Tekh. Fiz. **15**(6), 30 (1989) [Sov. Tech. Phys. Lett. **15**(3), 219 (1989)].
[4] V. A. Pozdeev, Prikl. Mat. Mekh. No. **6**, 1055 (1991).
[5] V. A. Pozdeev, *Nonsteady-State Wave Fields in Regions with Moving Boundaries* [in Russian], Naukova Dumka, Kiev (1992), 242 pp.

Translated by R. M. Durham

Transfer matrix method for media with quadratic optical nonlinearity

S. V. Fedorov

Institute of Laser Physics, 199034 St. Petersburg, Russia

M. A. Kaliteevskiĭ, N. V. Lukovskaya, and V. V. Nikolaev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted April 2, 1998)
Zh. Tekh. Fiz. **69**, 116–117 (April 1999)

A transfer matrix method is developed for media containing nonlinear layers. This method can be used to calculate the propagation of light, the electromagnetic field profile, and also second-harmonic generation in layered structures containing a layer exhibiting quadratic nonlinearity. © *1999 American Institute of Physics.* [S1063-7842(99)02204-7]

The transfer matrix method[1–5] is widely used to calculate the optical properties of layered structures. For a layered medium, problems involving the propagation of light in a layered structure and determining the frequencies of the natural optical[1] and polariton[3] modes of this structure are reduced to multiplying the transfer matrices across the individual layers forming the structure. For a linear medium the transfer matrices have dimensions of 2×2. The specific form of the matrices is determined by selecting a pair of parameters (basis) describing the electromagnetic field. The components of the electric and magnetic fields[2] tangential to the interface of the two media or the amplitudes of waves propagating in opposite directions[3] are most commonly used as the basis.

When various waves undergo mixing at some particular layer (for example, at a lattice of quantum wires, where the incident and diffracted waves are mixed), the dimensions of the transfer matrix are twice the number of waves being mixed.[4]

Recent advances in the field of vertical lasers, especially the experimental observation of second-harmonic generation in these systems,[6,7] necessitate the development of an adequate theoretical calculation method which can effectively solve the problem of light propagation in layered structures containing nonlinear layers.

The aim of the present paper is to develop a transfer matrix method for media with quadratic nonlinearity.

We shall consider a layered structure on which a wave of frequency ω and amplitude $\tilde{E}_1$ is incident. Let us suppose that the structure contains layers in which the light undergoes frequency doubling. The amplitude coefficients of reflection and transmission of light at frequencies ω and 2ω, denoted r_1, r_2 and t_1, t_2, may be obtained by solving the system of linear equations

$$\begin{pmatrix} t_1\tilde{E}_1 \\ 0 \\ t_2\tilde{E}_1 \\ 0 \end{pmatrix} = \tilde{E}\begin{pmatrix} \tilde{E}_1 \\ r_1\tilde{E}_1 \\ 0 \\ r_2\tilde{E}_1 \end{pmatrix}. \tag{1}$$

Here the transfer matrix across the structure $\hat{T}$ is the product of the transfer matrices across the various layers and the boundaries between them. The transfer matrix across a single nonlinear layer is determined by solving the system of differential equations[7]

$$\frac{dE_{\pm1}}{dz} = \pm ik_1E_{\pm1} - i\eta E^*_{\pm1}E_{\pm2}, \tag{2a}$$

$$\frac{dE_{\pm2}}{dz} = \pm ik_2E_{\pm2} - i\eta E^2_{\pm1}, \tag{2b}$$

where $E_{\pm1,2}$ are the amplitudes of waves propagating along the normal to the layers (the direction of the z axis) and in the opposite direction, respectively; $\eta(\omega)=(\omega^2/2c^2)\times\varepsilon_{i,kl}(\omega)e_ie_ke_l$ is the nonlinearity parameter, which is proportional to the sum of the components of the second-order nonlinear susceptibility tensor $\varepsilon_{ikl}(\omega, 2\omega)$.

If the nonlinearity is so small that it hardly changes the amplitude per pass through the nonlinear layer, the solution of the system (2) can be written as

$$E_{\pm1}(z) = \exp(\pm ik_1z)(A_{\pm1}E_{\pm10} - i\eta zE^*_{\pm10}E_{\pm20}), \tag{3a}$$

$$E_{\pm2}(z) = \exp(\pm ik_2z)(A_{\pm2}E_{\pm20} - i\eta zE^2_{\pm10}), \tag{3b}$$

where $E_{\pm10}$ and $E_{\pm20}$ are the wave amplitudes at the boundary of the nonlinear layer and $A_{\pm1}=1$ for $|\eta zE_{\pm10}|\ll1$.

In this approximation the transfer matrix across a nonlinear layer of thickness d_j will have the nondiagonal form

$$\hat{N}_j=\begin{pmatrix} A_{+1}e^{ik_1d_j} & 0 & -i\eta d_jE^*_{+10}e^{ik_1d_j} & 0 \\ 0 & A_{-1}e^{-ik_1d_j} & 0 & -i\eta d_jE^*_{-10}e^{-ik_1d_j} \\ -i\eta d_jE_{+10}e^{ik_2d_j} & 0 & A_{+2}e^{ik_2d_j} & 0 \\ 0 & -i\eta d_jE_{-10}e^{-ik_2d_j} & 0 & A_{-2}e^{-ik_2d_j} \end{pmatrix}. \tag{4}$$

For an arbitrary nonlinearity per pass through an isolated layer, the exact solution of the system (2), after two integrals of the equations are taken, is reduced to an elliptic integral.[8] This solution may be represented by formulas (3), assuming that the values of $A_{\pm 1,2}$ depend on the coordinate and the input amplitudes $A_{\pm 1,2}(z,E_{\pm 10},E_{\pm 20})$. In this case, the structure of the transfer matrix (5) remains the same, since the nonlinear wave interaction takes place in pairs.

In order to find the transfer matrix across the nonlinear layer (4), that is to say the transfer matrix of the entire system $\hat{T}$, we need to define the amplitudes of the ω and 2ω waves at the boundary of the nonlinear layer $E_{\pm 10}$ and $E_{\pm 20}$ in terms of the amplitude of the incident radiation $\tilde{E}_1$, but this can only be accomplished if we know the amplitude of the waves reflected from the system $r_{1,2}\tilde{E}_1$, since

$$\begin{pmatrix} E_{+10} \\ E_{-10} \\ E_{+20} \\ E_{-20} \end{pmatrix}=\hat{L}\begin{pmatrix} \tilde{E}_1 \\ r_1\tilde{E}_1 \\ 0 \\ r_2\tilde{E}_1 \end{pmatrix}, \tag{5}$$

where $\hat{L}$ is the transfer matrix from the edge of the structure to the boundary of the nonlinear layer.

However, the reflection coefficients $r_{1,2}$ are determined by solving Eq. (1), for which we need to know the transfer matrix $\hat{T}$ of the entire system.

We shall use perturbation theory in terms of the small nonlinearity parameter η to resolve this dilemma. In the zeroth approximation (i.e., for the linear theory) the transfer matrix in a nonlinear layer degenerates into a diagonal matrix, and the matrix of the entire system $\hat{T}^{(0)}$ and the coefficients of reflection $r^{(0)}_{1,2}$ and transmission $t^{(0)}_{1,2}$ do not depend on the amplitudes $E_{\pm 10}$ and $E_{\pm 20}$.

Having determined the form of the matrix of the entire system $\hat{T}^{(n)}$ in the nth order of perturbation theory, we can determine $r^{(n)}_{1,2}$ and $t^{(n)}_{1,2}$ using Eq. (1). We then need to determine the amplitudes $E^{(n+1)}_{\pm 10}$ and $E^{(n+1)}_{\pm 20}$ as follows:

$$\begin{pmatrix} E^{(n+1)}_{+10} \\ E^{(n+1)}_{-10} \\ E^{(n+1)}_{+20} \\ E^{(n+1)}_{-20} \end{pmatrix}=\hat{L}\begin{pmatrix} \tilde{E}_1 \\ r^{(n)}_1\tilde{E}_1 \\ 0 \\ r^{(n)}_2\tilde{E}_1 \end{pmatrix}. \tag{6}$$

Substituting Eq. (6) into Eq. (4), we can determine the matrices of the nonlinear layers $\tilde{N}^{(n+1)}_j$ and thus the matrices of the entire system $\hat{T}^{(n+1)}$ in the $(n+1)$th order of perturbation theory. Convergence of the iteration process for $n\to\infty$ is guaranteed[8] for $\eta D|\tilde{E}_1|<1$ where D is the total thickness of the structure. As a result, we solve a system of algebraic equations of the type (1) for the unknown amplitudes of the field in the nonlinear layers.

To conclude, we note that the proposed algorithm for the transfer matrices and spectral parameters of nonlinear layered systems is universal. It can easily be generalized to matrices of large dimensions when the radiation field is described by a large number of pairwise mixing plane waves and to a different type of nonlinearity of the layers.

The authors would like to thank the Russian Fund for Fundamental Research (Grant No. 97-02-18341).

[1] M. A. Kaliteevskiĭ and A. V. Kavokin, Fiz. Tverd. Tela. (St. Petersburg) **37**, 2721 (1995) [Phys. Solid State **37**, 1497 (1995)].
[2] M. Born and E. Wolf, *Principles of Optics*, 4th ed. [Pergamon Press, Oxford, 1969; Nauka, Moscow, 1970, p. 77].
[3] A. V. Kavokin and M. A. Kaliteevski, Solid State Commun. **95**, 859 (1995).
[4] A. V. Kavokin, M. A. Kaliteevski, and M. R. Vladimirova, Phys. Rev. B **54**, 1490 (1996).
[5] E. L. Ivchenko, M. A. Kaliteevski, A. V. Kavokin, and A. I. Nesvizhskii, J. Opt. Soc. Am. B **13**, 1061 (1996).
[6] N. Yamada, Y. Kaneka, S. Nakagama *et al.*, Appl. Phys. Lett. **68**, 1895 (1996).
[7] E. Rosencher, B. Vinter, and V. Berger, J. Appl. Phys. **78**, 6042 (1995).
[8] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, [Pergamon Press, Oxford, 1984; Nauka, Moscow, 1982, pp. 524–530.]

Translated by R. M. Durham

Transition to chaotization and loss of self-averagability in two-dimensional two-phase media at the percolation threshold

S. P. Luk'yanets

Institute of Physics, National Academy of Sciences of Ukraine, 252650 Kiev, Ukraine

A. E. Morozovskiĭ and A. A. Snarkskiĭ

Ukrainian National Technical University, 252056 Kiev, Ukraine

(Submitted April 21, 1998)
Zh. Tekh. Fiz. **69**, 118–121 (April 1999)

Giant fluctuations of the electric field, recently identified experimentally in two-dimensional, two-phase media at the percolation threshold, are discussed. A hierarchical realization of these media is used to show that for $\mathrm{Re}\,\sigma_i=0$ (where σ_1 and σ_2 are the conductances of the phases) and $\mathrm{Im}\,\sigma_2/\mathrm{Im}\,\sigma_1>0$ the hierarchy construction procedure yields the Dykhne expression $\sigma_e=\sqrt{\sigma_1\sigma_2}$ whereas for $\mathrm{Im}\,\sigma_2/\mathrm{Im}\,\sigma_1<0$ the procedure becomes chaotized and the medium loses its property of self-averagability. © *1999 American Institute of Physics.*
[S1063-7842(99)02304-1]

The effective conductance σ_e of a randomly inhomogeneous medium, which by definition relates the volume-averaged fields and currents, is a quantity which characterizes the overall conductance of an inhomogeneous sample. For a quasisteady-state field, σ_e is broken down into two terms, each having its own particular dependence on the field frequency, the conductances and permittivities of the phases, and their concentration. In two-phase, highly inhomogeneous media these dependences are steepest in the so-called critical region, where the concentration of the highly conducting phase is near the percolation threshold

Dykhne[1] derived an expression for the effective conductance σ_e of a two-dimensional two-phase medium with a geometric average equivalent phase configuration

$$\sigma_e=\sqrt{\sigma_1\sigma_2}, \tag{1}$$

where σ_i are the local conductance of the phases, at a concentration $p=0.5$.

In particular, these media (Dykhne media) include those at the percolation threshold. Various authors[2–5] have recently shown that for $\mathrm{Re}\,\sigma_1=0$ and $\mathrm{Re}\,\sigma_2\ll\mathrm{Im}\,\sigma_2$ these media exhibit giant fluctuations of the local electric field. In particular, this implies that the modulus of the electric field is not averaged over dimensions of the order of the standard correlation length (see the experiment described by Lagarkov *et al.*[4]). Éntin and Éntin[6] put forward the idea that a renormalization group mapping for complex impedances with low local losses leads to a dynamical chaos scenario and a fractal dependence of the impedance for hierarchical chains.

An example of a hierarchical construction of a medium with a geometrically equivalent phase configuration was described by Schulgaser,[7] Morozovskiĭ and Snarskiĭ.[8] Adopting the approach from Ref. 8 and considering the limiting case $\mathrm{Re}\,\sigma_2=\mathrm{Re}\,\sigma_1=0$, we can show that depending on the sign of $h=\mathrm{Im}\,\sigma_2/\mathrm{Im}\,\sigma_1$, the medium will have fundamentally different properties. For $h>h_c=0$, the medium is self-

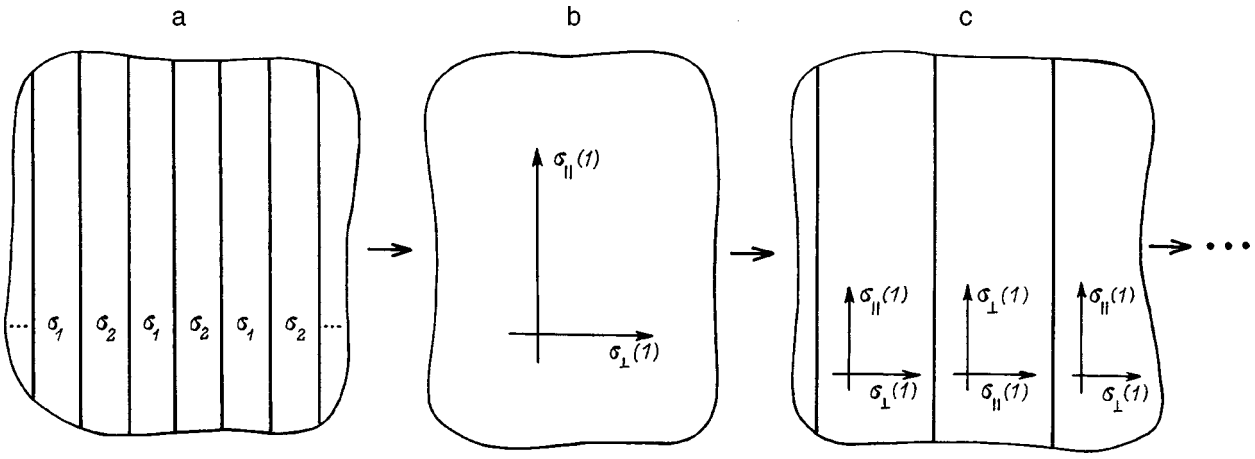

FIG. 1. Procedure for systematic construction of a Dykhne hierarchical medium.

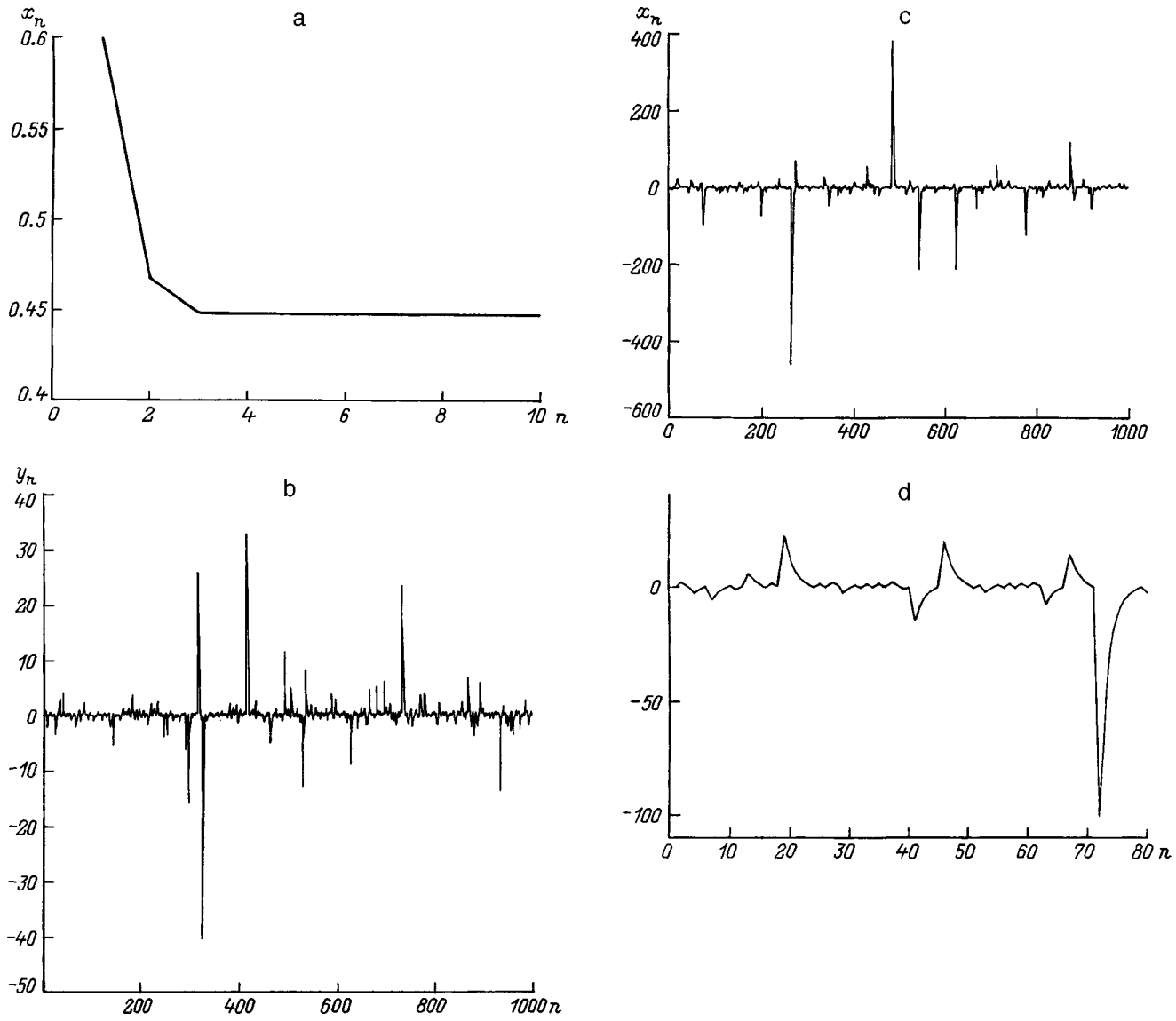

FIG. 2. a — Behavior of the iteration sequence (6) when the purely imaginary values of the conductances of the phases have the same sign, $h=\sigma_2/\sigma_1$; b–d — chaotic behavior of the iteration procedure for $h<0$: $h=-0.2$ (b) and -2 (c,d).

averaging, and the effective conductance is determined (as it should be) by the Dykhne formula (1). When $h<h_e=0$, the medium loses its self-averaging property, and it is difficult to introduce effective transport coefficients.

According to Ref. 8, at the first step the medium ''is assembled'' from strips of different thickness having the conductances σ_1 and σ_2 (Fig. 1a). By making the strip thicknesses tend to zero, i.e., by homogenizing the medium, we obtain a single crystal with the principal components of the conductance tensor $\sigma_\parallel(1)$ and $\sigma_\perp(1)$ (Fig. 1b). Then, by cutting strips of equal thickness from this crystal in the parallel and perpendicular directions, we assemble a new single crystal (Fig. 1c) for which the principal components of the conductance tensor are $\sigma_\parallel(2)$ and $\sigma_\perp(2)$. It is easy to show that at each stage we have

$$\sigma_\perp(n+1)=(\sigma_\perp(n)+\sigma_\parallel(n))/2,$$

$$\sigma_\perp(n+1)=2\sigma_\perp(n)\sigma_\parallel(n)/(\sigma_\perp(n)+\sigma_\parallel(n)). \tag{2}$$

Note that the numerical coefficients in Eq. (2) appear as a result of normalizing to the geometric dimensions of the total conductance of a medium with a large number (in the limit an infinite number) of strips. The iteration procedure (2) has the invariant

$$\sigma_\perp(n)\sigma_\parallel(n)=\sigma_1\sigma_2=a, \tag{3}$$

allowing for which Eq. (2) may be written in the form

$$\sigma_\perp(n+1)=(\sigma_\perp(n)+a/\sigma_\perp(n))/2,$$

$$\sigma_\parallel(n)=a/\sigma_\perp(n). \tag{4}$$

It is easy to show[8] that for cases of real $\sigma_1>0$ and $\sigma_2>0$, when $n\to\infty$ we obtain expression (1) for the effective conductance of the medium.[7] The mapping (4) does in fact have a fixed stable point $\sigma_\parallel(\infty)=\sigma_\perp(\infty)=\sigma_e=\sqrt{\sigma_1\sigma_2}$.

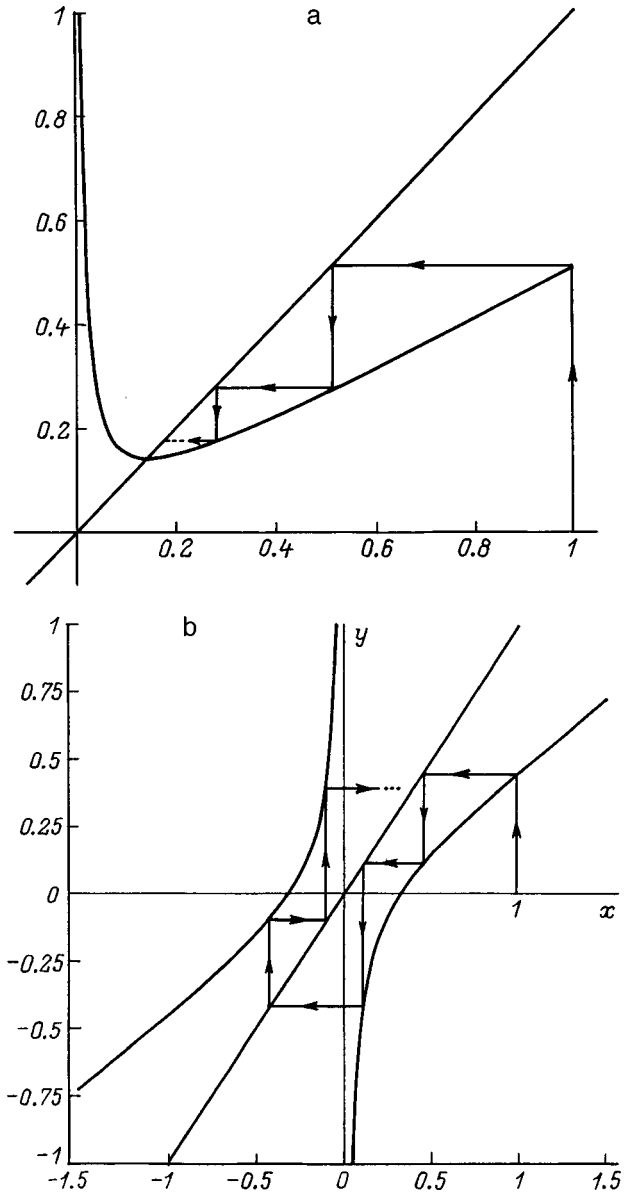

FIG. 3. Convergence of the iteration procedure with the mapping function $f(x)=(x+h/x)/2$ to the fixed stable point $\sqrt{h}$ for $h=0.02>0$ (a) and loss of stability of the iteration procedure for $h=-0.1<0$ (b).

For large n, $\sigma_\perp(n)$ has the form

$$\sigma_\perp(n)\approx\sqrt{\sigma_1\sigma_2}(1+0.5(\sigma_1/\sigma_2-1)\exp(-n)),$$

which suggests that the procedure converges rapidly (Fig. 2a).

We shall investigate Eq. (4) for purely imaginary values of σ_1 and σ_2 and we shall show that under certain conditions the procedure loses its fixed points and stable cycles, and the sequence $\sigma_\parallel(n)$ and $\sigma_\perp(n)$ becomes chaotized. This implies that the medium obtained by this procedure is not self-averaging. For purely imaginary σ_1 and σ_2 the elements of the sequence determined by the procedure (4) will also be imaginary $\mathrm{Re}\,\sigma_\perp(n)=0$, $\mathrm{Re}\,\sigma_\parallel(n)=0$. The iteration procedure has the same form as Eq. (4). Note that for purely imaginary $\sigma_\parallel$ and $\sigma_\perp$ the parameter a is real.

Two cases should be distinguished.

1) $a=\sigma_1\sigma_2<0$ (i.e., σ_1 and σ_2 have the same sign; for example, the elements of the first and second phases are capacitances whose active resistances can be neglected). In this case, we arrive at the same result as for real σ_1 and σ_2. The procedure has stable fixed points and the effective conductance is given by

$$\sigma_e=\sigma_\perp(\infty)=\sigma_\parallel(\infty)=\begin{cases} i\sqrt{|\sigma_1\sigma_2|}, & \mathrm{Im}\,\sigma_1>0, \quad \mathrm{Im}\,\sigma_2>0, \\ -i\sqrt{|\sigma_1\sigma_2|}, & \mathrm{Im}\,\sigma_1<0, \quad \mathrm{Im}\,\sigma_2<0. \end{cases} \tag{5}$$

2) $a=\sigma_1\sigma_2>0$ (i.e., σ_1 and σ_2 have different signs, for example, the elements of the first phase are capacitances and those of the second phase are inductances whose active resistances can be neglected). In this case, for the mapping given by Eq. (4) no fixed points or stable cycles exist. It follows directly from Eq. (4) that for $a=\sigma_1\sigma_2>0$, $\sigma_\parallel(n)$ and $\sigma_\perp(n)$ have different signs for any n. This implies that in this case, the medium does not become isotropized. For convenience we convert to the dimensionless variables $x_n=\mathrm{Im}\,\sigma_\perp(n)/|\sigma_1|$ and $y_n=\mathrm{Im}\,\sigma_\parallel(n)/|\sigma_1|$ in Eq. (4), and then this equation can be rewritten in the form

$$x_{n+1}=(x_n+h/x_n)/2, \quad y_n=h/x_n, \tag{6}$$

where $h=\sigma_2/\sigma_1$. The initial value is now $x_0=\pm1$ and the parameter defining the different media is h. Figure 2 shows the behavior of the iteration seqence (6) for various h. When $h>0$ the sequence x_n converges to a stable fixed point x_∞ (Fig. 2a). When h becomes negative, the sequence ceases to converge and as h decreases, it increasingly exhibits traits of chaotic behavior (Fig. 2b–2d). The procedure (6) in fact leads to chaotic dynamics, and for the mapping (4) $f(x)=(x^2+a)/2x$ (where a is positive and real) the Julia set J_f separating the basins of attraction of the stable fixed points $\pm\sqrt{h}$ coincides with the imaginary axis.[9] On J_f the mapping (4) induces a one-dimensional mapping which reduces to Eq. (6) and determines the dynamics on the Julia set. The mapping (4) is conjugate with the mapping $R(u)=u^2$ obtained by substituting $u=(x+\sqrt{h})/(x-\sqrt{h})$. The imaginary axis (Julia set) then becomes a single circle on which the dynamics are defined by the mapping $r(\theta)=2\theta(\mathrm{mod}\,1)$. As we know,[9,10] this mapping generates chaotic dynamics.

The qualitative behavior of x_n as a function of n may be explained using a graph of the mapping function $f(x)=(x+h/x)/2$ which determines the iteration procedure (Fig. 3). Figure 3a shows a sequence of iterations and its convergence to the fixed stable point $\sqrt{h}$ for $h>0$. For fairly small negative h (Fig. 3b) the sequence x_n initially decreases monotonically to zero ($f(x)\approx x/2$, $|h|\ll|x|$), although the map has a ''dip'' near zero ($f(x)\approx-|h|/2x$, $|x|\ll|h|\ll1$), as a result of which the sequence loses its monotonicity, changes sign, and after a certain number of iterations, returns to the region of high values (compared with $|h|$) and again begins to tend monotonically to zero. As h decreases further, the dip near the zero increases, which reduces the intervals of monotonicity, and increases the irregular regions and the amplitudes x_n.

This procedure can also be used to obtain the spectrum of resonant frequencies which is determined by the set of values $h=\sigma_2/\sigma_1$ whose number increases with the iteration step n as 2^n, for which the procedure (6) diverges.

To sum up, an example of one possible realization of Dykhne media with a hierarchical structure has been used to show that for certain values of the conductances of the phases, a medium, while remaining a Dykhne medium in the geometrical sense (two-dimensional, two-phase, geometrically equivalent phase configuration), loses its self-averagability in terms of physical properties, and the concept of effective transport coefficients is thereby lost. Topics for further discussion are whether other possible procedures resulting in Dykhne media lose their property of self-averagability and whether the ''usual'' checkerboard realization of a Dykhne medium loses its self-averagability.

We are grateful to É. M. Baskin and M. V. Éntin for discussions on the absence of self-averagability of the effective properties of weakly absorbing media, and to M. V. Éntin for allowing us to familiarize ourselves with Ref. 6 prior to publication and for fruitful discussions of this work which improved our understanding of the issues raised.

This work was partially supported by Grants from the Russian Fund for Fundamental Research No. 97-02-18397 and ''Nauka-Servis'' No. 1-200-96.

[1] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1970)].
[2] F. Brauers, S. Blacher, and A. K Sarychev, *Fractal Reviews in the Natural and Applied Sciences*, edited by M. Novak (Chapman and Hall, 1995).
[3] F. Brauers, S. Blacher, N. Henrioulle, and A. K Sarychev, in *Book of Abstracts of the Fourth International Conference on Electrical Transport and Optical Properties of Inhomogeneous Media*, *ETOPIM4*, Moscow, 1996. p. 46 [Submitted to J. Phys. A].
[4] A. N. Lagarkov, K. V. Rozanov, A. K Sarychev, and N. A Simonov, *ibid.* p. 56.
[5] E. M. Baskin, M. V. Entin, A. K Sarychev, and A. A Snarskii, Physica A **242**, 49 (1997).
[6] M. V. Éntin and G. M. Éntin, JETP Lett. **64**, 467 (1996).
[7] K. J Schulgaser, J. Phys. C **10**, 407 (1977).
[8] A. E. Morozovskiĭ and A. A Snarskiĭ, Ukr. Fiz. Zh. **28**, 1230 (1983).
[9] H.-O. Peitgen, *The Beauty of Fractals: Images of Complex Dynamical Systems* [Springer-Verlag, New York, 1986; Mir, Moscow, 1993, p. 93].
[10] H. G. Schuster, *Deterministic Chaos* [Physik-Verlag, Weinheim, 1984; Mir, Moscow, 1988].

Translated by R. M. Durham

Thermomechanical effect in a planar nematic induced by a quasistatic electric field

R. S. Akopyan, R. B. Alaverdyan, S. Ts. Nersisyan, É. A. Santrosyan, and Yu. S. Chilingaryan

Erevan State University, 375049 Erevan, Armenia

(Submitted May 15, 1998)
Zh. Tekh. Fiz. **69**, 122–124 (April 1999)

A thermomechanical flow of uniformly oriented nematic liquid crystal induced by a quasistatic electric field is observed. This flow occurs when the electric field strength exceeds the static Fréedericksz transition threshold. The effect is attributed to an electric-field-induced nonuniformity of the director orientation which is required for the onset of the thermomechanical effect. The experimental results show good agreement with the theoretical estimates.
 [S1063-7842(99)02404-6]

1. Thermomechanical effects in cholesteric liquid crystals have been studied both theoretically and experimentally.[1–4] Akopyan and Zel'dovich[5] predicted thermomechanical effects in nematic liquid crystals (NLCs) with a nonuniform director distribution. In particular, those authors showed that a prerequisite for the establishment of a thermomechanical hydrodynamic flow is the presence of three gradients: temperature, director distribution, and hydrodynamic flow velocity. The thermomechanical rotation of an NLC film was first observed experimentally by Lavrentovich and Nastishin,[6] who measured the thermomechanical coefficient. Akopyan *et al.*[7] observed a thermomechanical effect in a hybrid-oriented NLC and also investigated oscillatory regimes for the thermodynamic flow velocity. It was demonstrated experimentally that no thermomechanical effect occurs in homogeneous planar and homeotropically oriented NLCs.

Here we report the observation of a thermomechanical hydrodynamic flow in a planar oriented NLC induced by a quasistatic electric field. The effect is attributed to a nonuniform reorientation of the NLC director above the Fréedericksz transition threshold. However, the hydrodynamic flow occurs in the presence of a vertical temperature gradient and a nonuniform director distribution. This effect provides valuable information on the electroelastic properties of NLCs.

2. For the experiments we used a cell containing a planar oriented sample of the nematic liquid crystal 5CB (Fig. 1). At the bottom of the substrate a planar boundary condition *1* was imposed by polishing. The boundary condition for the top of the substrate is free. As a result we have a uniform planar orientation for the NLC director. We used glass substrates *3*, 1 mm thick, on which we deposited electrically conducting coatings. A quasistatic electric field was applied to the substrates. A vertical temperature gradient was created by using circulating water *4* with two controllable temperatures. The thermal conductivity of the substrates was between two and three times that of the NLC film. This is an important condition to ensure that the boundary conditions are constant for the temperature. The hydrodynamic motion was observed by adding aluminum powder *2* to the NLC at a concentration of the order of 10^{-3} wt %. The temperature difference was varied between 0 and 10 °C with an error of 0.1 °C.

In the absence of the quasistatic electric field we detected no thermomechanical effect involving the appearance of a hydrodynamic flow perpendicular to the temperature gradient. For a cell of thickness $L=70\ \mu$m and a temperature gradient $\Delta T=4.7\,^{\circ}$C the threshold electric field for the appearance of a hydrodynamic flow was $U_{\text{thr}}=0.25$ V. As the strength increased, the flow velocity increased, tending to saturate (Fig. 2a). This increase was caused by an increase in the orientation gradient of the NLC director. At $U=14$ V the velocity was $V_{\max}\approx 1.06\ \mu$m/s. As the field was increased further, the director orientation gradient and the thermome-

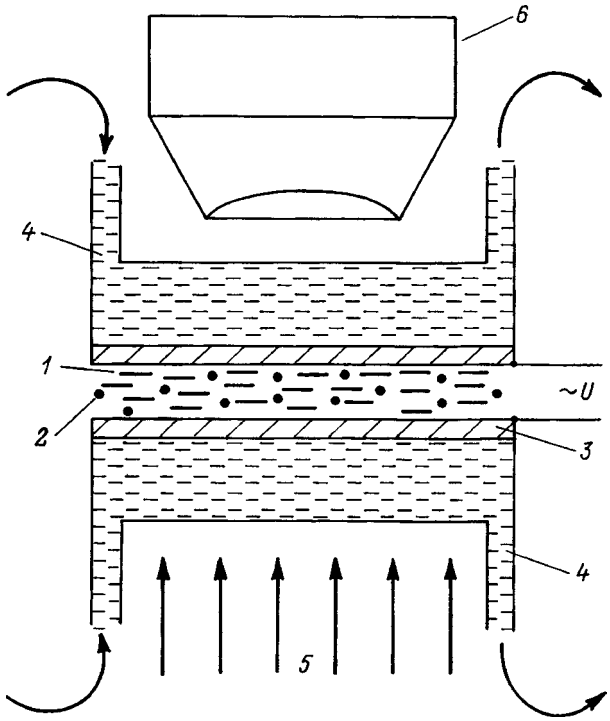

FIG. 1. *1* — Planar oriented nematic liquid crystal, *2* — particles of aluminum powder, *3* — glass substrates with transparent electrically conducting coating, *4* — circulating water, *5* — optical illumination, and *6* — microscope.

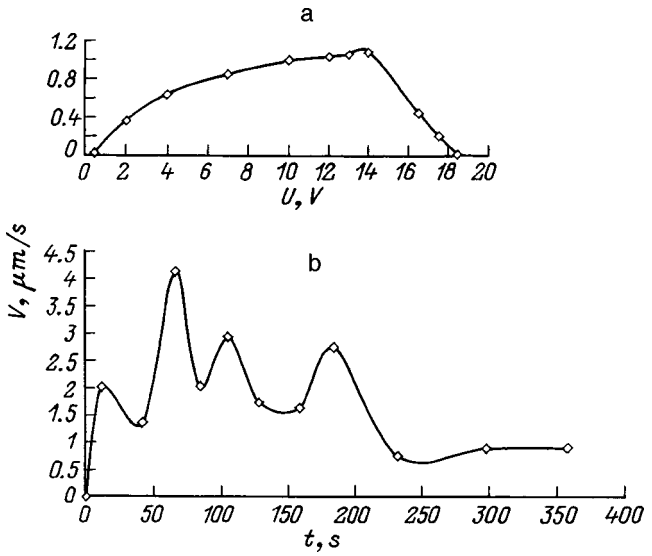

FIG. 2. Electric field strength U versus liquid crystal flow velocity V for a cell thickness $L=70\ \mu$m and temperature gradient $\Delta T=4.7°$ (a) and oscillating behavior of the establishment of the hydrodynamic flow velocity (b).

chanical hydrodynamic flow velocity fell to zero at $U=18.4$ V. Before a steady-state flow was established, the velocity exhibited oscillatory behavior caused by competition between reorientations of the NLC director and the hydrodynamic flow (Fig. 2b).

3. For the theoretical estimates we shall first consider the thermomechanical effect in a planar oriented nematic with the director **n**: $n_x=\sin\theta$, $n_y=0$, $n_z=\cos\theta$. For instance, for the hybrid orientation we have $\theta=\pm(\pi/2L)$. Let us assume that the temperatures $T(z=0)>T(z=L)$ are maintained on the substrates of the liquid crystal cell. Then, if a director orientation gradient exists along the z coordinate, a thermomechanical hydrodynamic flow appears in the direction of the x axis: $\mathbf{V}(\mathbf{z})=\mathbf{e}_x V_x(z)$. We shall consider a situation which is steady-state ($\partial/\partial t=0$) and homogeneous in the (x,y) plane ($\partial/\partial x=\partial/\partial y=0$). In the thermomechanical single-constant approximation ($\xi_1=\xi_2=\ldots\xi_{12}=\xi$) we obtain the Navier–Stokes equation in the form[5]

$$(\eta_1 n_x^2+\eta_2 n_z^2)\frac{d^2V_x}{dz^2}+2(\eta_1-\eta_2)n_x n_z\frac{d\theta}{dz}\frac{dV_x}{dz}$$
$$=\frac{1}{4}\xi\nabla T\left(\frac{d\theta}{dz}\right)^2(3-n_x n_z-15n_x^2+16n_x^3 n_z+12n_x^4)$$
$$+\frac{1}{4}\xi\nabla T\frac{d\theta}{dz}(3+3n_x n_z-n_x^2-3n_x^3 n_z+4n_x^4). \quad (1)$$

Here $\nabla T=(dT/dz)$ is the temperature gradient and η_1 and η_2 are the Miesowicz viscosity coefficients. For the specific form of the distribution $\theta(Z)$ we need to solve Eq. (1) numerically with the boundary conditions $V_x(z=0)=V_x(z=L)=0$ and obtain the velocity distribution $V_x(z)$. This distribution is close to parabolic with a maximum at the center of the cell. For a hybrid cell this maximum has the form

$$V_{\max}\approx 0.148\frac{\Delta T}{L(\eta_2-\eta_1)}. \quad (2)$$

We shall now analyze a nematic liquid crystal with an initially uniform director distribution. No thermomechanical hydrodynamic flow appears in this cell in the absence of a director gradient. However, under the action of external fields the director distribution may become distorted and induce a thermomechanical effect. We know[8] that at the threshold electric field

$$E_{\text{thr}}=\frac{\pi}{2L}\sqrt{\frac{4\pi K}{\varepsilon_a}},$$

where K is the Frank elastic constant and ε_a is the anisotropy of the permittivity, a Fréedericksz transition occurs wherein the molecules are reoriented by the angle θ:

$$\theta=\theta_m\sin\frac{\pi z}{2L},\quad \theta_m=\sqrt{\frac{(E-E_{\text{thr}})}{E_{\text{thr}}}}. \quad (3)$$

Substituting the distribution (3) into Eq. (1) and integrating, we obtain the maximum velocity

$$V_{\max}=\frac{\pi\Delta T}{64L\eta_2}\xi\theta_m. \quad (4)$$

This dependence of the hydrodynamic flow velocity on the temperature gradient $(\Delta T)/L$ shows good agreement with the experimental results. Thus, in an NLC cell with an initially uniform director distribution a field may induce a thermomechanical effect whereby, in the presence of a vertical temperature gradient, a uniform hydrodynamic flow is established perpendicular to the normal of the plane of the field-induce d reorientation of the director and the temperature gradient. In these experiments the nonuniformity of the director distribution was provided by a quasistatic electric field. However, in uniform cells a thermomechanical effect can be induced by static magnetic and optical fields.

[1] F. M. Leslie, Proc. R. Soc. London, Ser. A **307**, 359 (1968).
[2] M. J. Stephen and J. P. Straley, Rev. Mod. Phys. **46**, 617 (1974).
[3] J. Janossy, Mol. Cryst. Liq. Cryst. Lett. **72**, 233 (1982).
[4] V. F. Shabanov, S. Ya. Vetrov, G. M. Zharkova *et al.*, Opt. Spektrosk. **63**, 1368 (1987) [Opt. Spectrosc. **63**, 811 (1987)].
[5] R. S. Akopyan and B. Ya. Zel'dovich, Zh. Éksp. Teor. Fiz. **87**, 1660 (1984) [Sov. Phys. JETP **60**, 953 (1984)].
[6] O. D. Lavrentovich and Yu. A. Nastishin, Ukr. Fiz. Zh. **32**, 710 (1987).
[7] R. S. Akopyan, R. B. Alaverdyan, É. A. Santrosyan *et al.*, Pis'ma Zh. Tekh. Fiz. **23**(17), 77 (1997) [Tech. Phys. Lett. **23**(9), 690 (1997)].
[8] L. M. Blinov, *Electro-Optical and Magneto-Optical Properties of Liquid Crystals* [Wiley, New York, 1983; Nauka, Moscow, 1978].

Translated by R. M. Durham

Heat release accompanying current injection into a composite superconductor in a stable superconducting state

V. R. Romanovskiĭ

Kurchatov Institute Russian Science Center, 123182 Moscow, Russia
(Submitted May 15, 1998)
Zh. Tekh. Fiz. **69**, 125–128 (April 1999)

An investigation is made of the Joule dissipation of energy in the superconducting phase of a composite superconductor as the current in it increases. It is shown that at high rates of current injection the dissipative processes are characterized by an appreciable release of heat, which substantially influences the critical energies for the loss of superconductivity and the velocity of the normal zone. An analytical expression is given for calculating the power of the thermal losses during current injection.
[S1063-7842(99)02504-0]

Studies of thermal losses in composite superconductors form the basis for estimating the stability of their thermal properties. In thermal stabilization theory it is generally assumed that the current flowing inside the composite varies negligibly.[1,2] Consequently, in descriptions of the thermal processes taking place in superconducting windings the Joule heat release is neglected at temperatures between the coolant temperature and the so-called resistive transition temperature. In experimental investigations of the kinetics of the normal zone in superconducting composites carrying a rapidly varying current, Iwakuma *et al.*[3] and Vysotsky *et al.*[4] observed anomalously high normal-zone propagation velocities whose existence is outside the limits of the theory developed so far. In order to explain these effects, Pukhov and Rakhmanov[5,6] proposed a model which assumes that the heating of a superconducting composite depends weakly on the details of the temperature dependence of the heat release in the superconducting state and is a consequence of a reduction in the current-carrying capacity of the superconductor caused by an increase in the rate of current rise.

We shall solve this problem in a nonisothermal approximation assuming that the superconductor temperature changes in response to the action of any external perturbing factor. This allows us to study the laws governing the loss of superconductivity in a more correct formalism.[7,8]

Let us assume that initially no current flows in a cooled composite superconductor of circular cross section and that the current then begins to increase linearly at a given rate. We shall determine the instantaneous distributions of the temperature and electromagnetic field inside a composite situated in its self-magnetic field by solving the system of equations

$$c\frac{\partial T}{\partial t}=\frac{1}{r}\frac{\partial}{\partial r}\left(\lambda r\frac{\partial T}{\partial r}\right)+EJ, \quad \mu_0\frac{\partial J}{\partial t}=\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E}{\partial r}\right),$$

$$J=\eta J_s+(1-\eta)J_m,$$

$$E=J_s\rho_s\exp\left(\frac{J_s}{J_\delta}+\frac{T-T_C}{T_\delta}\right)=J_m\rho_m \tag{1}$$

with the initial and boundary conditions

$$T(r,0)=T_0, \quad E(r,0)=0,$$

$$\lambda\frac{\partial T}{\partial t}+h(T-T_0)\Big|_{r=r_0}=0, \quad \frac{\partial E}{\partial r}\Big|_{r=r_0}=\frac{\mu_0}{2\pi r_0}\frac{dI}{dt}.$$

Here c and λ are the specific heat per unit volume and the thermal conductivity of the composite in the transverse cross section, ρ_s and ρ_m are the electrical resistivities of the superconductor and the matrix, J_s and J_m are the currents flowing in the superconductor and the matrix, respectively, η is the coefficient of filling of the composite by the superconductor, h is the coefficient of heat transfer, T_0 is the coolant temperature, T_C is the critical temperature of the superconductor, and J_δ and T_δ are the given rise parameters of the current–voltage characteristic of the superconductor.

In addition to the numerical model, we make use of the characteristic features of the conditions for stable current injection[9,10] to write a simplified analytical expression for calculating explicitly the thermal loss power averaged over the cross section of the composite. First we shall consider the range of currents whose upper limit is the so-called cutoff current I_m, above which the superconducting state is unstable. Second, we shall allow for the existence of a maximum current $I_{\max}$ flowing in a superconducting composite with a nonlinear current–voltage characteristic, assuming an infinitely slow rate of current injection.[11] Finally, we shall bear in mind that the temperature distribution over the composite cross section is almost uniform but the density of the current flowing in the matrix is much lower than the critical density of the superconductor. Subject to these assumptions, we approximate the averaged power of the Joule heat release by the formula

$$G=\frac{1}{S}\int_S EJ\,ds=\frac{\eta J_{C0}(1-\Delta)}{S}\int_S E\,ds.$$

Here J_{C0} is the critical current density of the superconductor at the coolant temperature, Δ is a parameter which takes into account the nonlinear nature of the superconductor current–

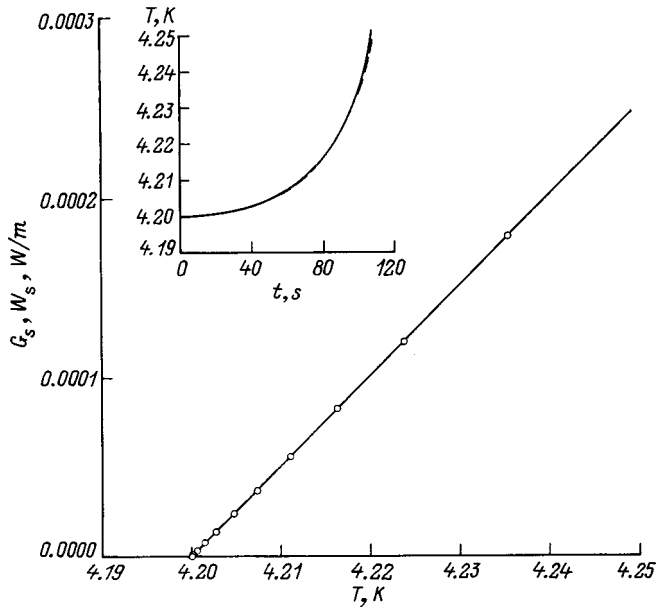

FIG. 1. Time variation of the Joule heat release and composite temperature for $dI/dt=10$ A/s.

voltage characteristic, and E is the electric field strength. For a wire of circular cross section this field is given by $E=(\mu_0/2\pi)(dI/dt)\ln(r/r_p)$, where $r_p(t)$ is the penetration depth of the electric field, which is related to the injected current by $I(t)=\pi\eta J_{C0}(1-\Delta_j)(r_0^2-r_p^2)$. Supplementing these expressions with the condition that in the limit $\lim_{r_p\to 0} I\to I_{\max}$ and performing suitable transformation, we write the desired expression for the heat release power in the superconducting phase

$$G(t)=\frac{\mu_0}{4\pi}\frac{dI}{dt}\frac{I_{\max}}{S}\left[\ln\frac{I_{\max}}{I_{\max}-I(t)}-\frac{I(t)}{I_{\max}}\right]. \tag{2}$$

This formula can be used to determine the background temperature of a superconducting composite preceding the onset of instability for arbitrary rates of injection. For all $I<I_m$ this temperature is a solution of the heat balance equation

$$c\frac{dT}{dt}=-\frac{hp}{S}(T-T_0)+G(t),\quad T(0)=T_0 \tag{3}$$

and we obtain in quadratures

$$T(t)=T_0+\frac{\mu_0}{4\pi}\frac{dI}{dt}\frac{I_{\max}}{cS}\int_0^t\left[\ln\frac{I_{\max}}{I_{\max}-I(\tau)}-\frac{I(\tau)}{I_{\max}}\right]\times\exp\left[-\frac{hp}{cS}(t-\tau)\right]d\tau. \tag{4}$$

Figures 1 and 2 give temperature dependences of the Joule losses released in a wire in a stable superconducting state per unit length $G_s=GS/2\pi$ for all typical rates of current injection. For the calculations the initial parameters were taken as: $r_0=5\times10^{-4}$ m, $c=1000$ J/(m$^3\cdot$K), $\lambda=100$ W/(mK), $h=10$ W/(m$^2\cdot$K), $T_0=4.2$ K, $\eta=0.5$, $\rho_s=5\times10^{-7}\,\Omega\cdot$m, $\rho_m=2\times10^{-10}\,\Omega\cdot$m, $J_\delta=4\times10^7$ A/m^2,

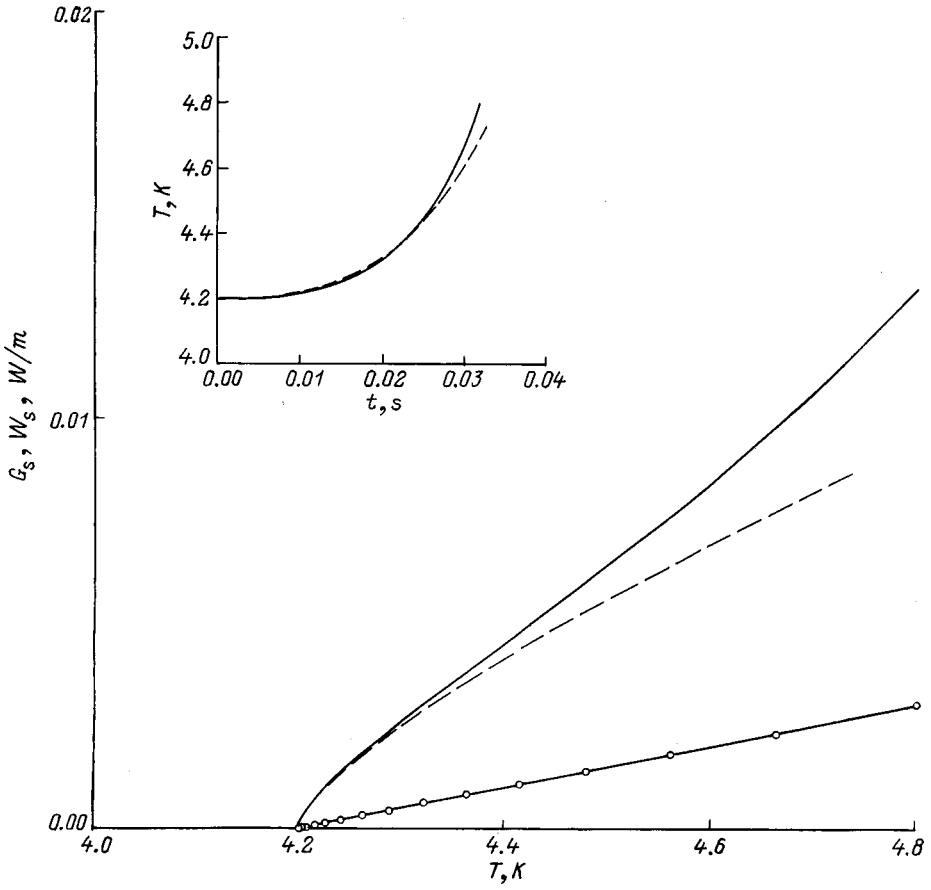

FIG. 2. Time variation of the Joule heat release and composite temperature for $dI/dt=10^4$ A/s.

$T_\delta = 0.048$ K, and $T_{CB} = 9$ K, which describe the temperature-averaged thermophysical and electrophysical properties of a niobium titanium superconductor in a copper matrix cooled with liquid helium. For these parameters we obtain $I_{\max} = 1226$ A, and the injection cutoff current has the value $I_m = 1120$ A for $dI/dt = 10$ A/s and the value $I_m = 367$ A for $dI/dt = 10^4$ A/s.

The solid curves in Figs. 1 and 2 give the numerical calculations made by solving the system (1) while the dashed curves give those using formulas (2) and (4). The straight lines with data points show the change in the power of the heat flux transferred to the coolant, also per unit length of the composite $W_s = r_0 h(T - T_0)$. The insets compare the corresponding values of the composite background temperature calculated using the two models described above.

As expected, at slow rates of current injection the Joule heat release is negligibly higher than the heat flux transferred to the coolant over the entire range of superconductor background temperature all the way to the onset of instability. Therefore the electric field induced by the current injection barely heats the composite. As the rate of current injection increases, this difference increases. As a result, in the stable stage of current injection the Joule heat release may appreciably exceed the cooling power of the coolant and the corresponding power of the thermal losses obtained for slow injection. For instance, for $I = 200$ A, which is attained in the case $dI/dt = 10$ A/s, the heat release power is $G_s' = 0.302 \times 10^{-5}$ W/m, while for $dI/dt = 10^4$ A/s it is $G_s'' = 0.299 \times 10^{-2}$ W/m, i.e., three orders of magnitude higher. As can be deduced from the results, this difference increases as the current increases further. Consequently, at fast rates of injection the permissible increase in the composite temperature, against the background of which any instabilities can develop, becomes very appreciable.

Adequate allowance for the role of the temperature factor in describing the dynamics of the dissipative processes taking place in superconductors also forms the basis for explaining the difference between the results obtained by the numerical and analytical models. The largest difference is observed directly before the onset of instability. In this context, we should note that it is incorrect to use formula (2) to determine the heat released after the onset of instability. In this case, the magnetic flux diffusion inside the composite is an avalanche-like process which will be accompanied by increasingly intensive heat release. An analysis of these states is outside the scope of the simplifying assumptions used above which do not permit taking into account in an analytical model a substantia lly nonsteady-state magnetic flux penetration process initiated by instability.

Changes in the thermal state of the superconducting phase of a composite superconductor and the existence of an appreciable additional heat release must necessarily be taken into account in theoretical analyses of the limits of permissible thermal perturbations and the velocity of the normal zone in its irreversible propagation along the composite. We shall assess the possible qualitative influence of this effect using the approximate formulas given in Ref. 5. Then under conditions of poor cooling one can obtain the following estimates for the relative decrease in the critical energy $\Delta_e = e''/e'$ and the relative increase in the velocity of the normal zone $\Delta_V = V''/V'$ that will take place upon an r-fold increase in the power of heat release $(r = q''/q')$ for the same current injected at two different rates dI'/dt and dI''/dt:

$$\Delta_e \cong 1 - 1.5(r-1)q'/(1-i-q'),$$

$$\Delta_V \cong 1 + 0.5(r-1)q'/(1-i-q'),$$

where i is the dimensionless current and q is the dimensionless power of the additional heat release.

These expressions show that even when the injected current does not reach the cutoff current, an increase in the additional Joule heat release is accompanied by an abrupt increase in the decay rate Δ_e and the rise rate Δ_V. Consequently, even at the stable current injection stage the critical energies may decrease substantially and the velocity of the normal zone may increase.

To conclude, we draw attention to the fact that for all $dI/dt > 0$ the curves of G_s and W_s describing the temperature dependences of the released and extracted heat have only one common point at $T = T_0$. Physically this is attributable to the nonsteady-state behavior of the instantaneous composite temperature described by Eq. (3). Thus, at the stable current injection stage, the temperature field of the composite will only decay to some thermodynamically stable temperature distribution depending on the features of the superconductor current–voltage characteristic and its cooling rate when $dI/dt = 0$. In the existing theory this specific feature of the evolution of the thermal processes in composite superconductors with current injection is neglected.

It therefore follows that the intensity of the heat release in the superconducting phase depends on the change in current and, appreciable thermal losses occur at high rates of increase. The heat released in this case differs by several orders of magnitude from that released under slow injection. This heat release is accompanied by an appreciable rise in the composite temperature as a result of the nonsteady-state dissipative processes taking place in the superconductor. Results obtained by both numerical and analytical solutions were compared. The latter can provide both qualitative and quantitative descriptions of the energy dissipation process that occurs in real superconducting wires without destroying the superconductivity.

This work was supported by the Russian Fund for Fundamental Research (Project No. 97-02-16066a).

[1] V. A. Al'tov, V. B. Zenkevich, M. G. Kremlev, and V. V Sychev, *Stabilization of Superconducting Magnet Systems* [in Russian], Energoatomizdat, Moscow (1984), 312 pp.
[2] M. Wilson, *Superconducting Magnets* [Oxford University Press, London, 1983; Mir, Moscow, 1985, 407 pp.].
[3] M. Iwakuma, H. Kanetaka, K. Tasaki *et al.*, Cryogenics **30**, 686 (1990).
[4] V. S. Vysotsky, V. N. Tsikhon, and G. B. J. Mulder, IEEE Trans. Magn. **MAG-28**, 735 (1992).
[5] A. A. Pukhov and A. L. Rakhmanov, Sverkhprovodimost' (KIAE) **5**, 1620 (1992).
[6] A. A. Pukhov and A. L. Rakhmanov, Sverkhprovodimost' (KIAE) **6**, 1165 (1993).

[7] V. R. Romanovskiĭ, JETP Lett. **59**, 709 (1994).
[8] V. R. Romanovskiĭ, Dokl. Akad. Nauk SSSR **350**, 752 (1996) [Phys. Dokl. **41**, 460 (1996)].
[9] A. V. Gurevich R. G. Mints, and A. L. Rakhmanov, *Physics of Composite Superconductors* [in Russian], Nauka, Moscow (1987), 240 pp.
[10] V. E. Keilin and V. R. Romanovskii, Cryogenics **33**, 986 (1993).
[11] E. Yu. Klimenko, N. N. Martovetskiĭ, and S. I. Novikov, Dokl. Akad. Nauk SSSR **261**, 1350 (1981) [Sov. Phys. Dokl. **26**, 1180 (1981)].

Translated by R. M. Durham

Ellipsometry as a rapid method of establishing a correlation between the porosity and the gas sensitivity of tin dioxide layers

D. Ts. Dimitrov, V. V. Luchinin, V. A. Moshnikov, and M. V. Panov

St. Petersburg State Electrical Engineering University, 197376 St. Petersburg, Russia
(Submitted May 19, 1998)
Zh. Tekh. Fiz. **69**, 129–130 (April 1999)

It is shown for the first time that reflection ellipsometry may be an effective method of establishing a correlation between the main parameters of gas-sensitive layers and their optical characteristics, suitable for rapid monitoring. This can advance the capability of predicting the properties of a material from its preparation conditions and consequently can reduce the optimization cycle time and thus lower the cost of a technology. © *1999 American Institute of Physics.* [S1063-7842(99)02604-5]

Recently, increasing attention has been focused on the micro- and nanoporosity of materials, which substantially influences their properties as gas-sensitive detectors. In order to predict the gas-sensitive properties of materials used in adsorption detectors, it is relevant to develop rapid methods of monitoring the porosity of gas-sensitive layers to predict their sensitivity and selectivity.

The samples used for the present study are based on the conventional material for gas-sensitive sensors,[1] SnO_{2-x}, doped with highly volatile impurities (tellurium and iodine) which create vacancy defects and pores in the layers during the treatment process. The aim of the present paper is to demonstrate that reflection ellipsometry is an effective method of monitoring the layer parameters by establishing a correlation between their porosity, gas sensitivity, and optical parameters.[2]

A distinct disadvantage of SnO_{2-x} sensors is their relatively low selectivity for specific gases, which impedes detection in mixtures containing several reducing gases. We know that selectivity can be achieved by integrating several detectors having different relative sensitivities.[3] Altering the porosity of the material is also an effective method of controlling the sensitivity to different gases.[4] In the present case, gas-sensitive layers of tin dioxide supersaturated with vacancies and pores were formed by doping the initial tin charge with highly volatile impurities such as iodine and also iodine and tellurium. Layers of tin doped with impurities were prepared on 22KhS ceramic, glass ceramic, Polykor, and optical quartz substrates (the refractive indices of the substrates were determined by ellipsometry using a pure surface model) by explosive evaporation. The evaporation temperature was 1150 °C. The layers were then oxidized in a controlled rare gas (argon) atmosphere at oxygen partial pressures of 0.25 and 0.35 atm. The oxidation process consisted of two stages: low-temperature annealing at $T=225$ °C for 6 h followed by high-temperature annealing at $T=525$ °C for 30 h. The technological aspects are described in detail by Dolotov *et al.*[5] Bakin *et al.*[4] showed that iodine is removed from the samples as I_2, SnI_2, or SnI_4, forming a porous structure. During the treatment process the pore concentration depends on the initial level of doping. Table I gives the relative gas sensitivity of layers of tin dioxide for toluene at a concentration of 60 mg/m^3.

In the present paper we suggest using an optical rapid diagnostic technique based on ellipsometric measurements of the effective refractive index to establish a correlation between the technological parameters of the layer synthesis process and their gas sensitivity.

The ellipsometric angles Ψ and Δ for these layers were determined using an updated LÉF-2 device, and the refractive index and the thickness of the transparent layers of tin dioxide were calculated using the Holmes method.[6] The adequacy of the isotropic substrate–isotropic layer model is confirmed by the satisfactory agreement between the results

TABLE I. Characteristics of technological preparation conditions of tin dioxide layers.

Sample No.	Quantity andcomposition of dopant in initial mixture	Oxygen pressure during synthesis, atm	Sensitivity to toluene, %	Refractive index, n
1	Undoped SnO_2	0.25	56	1.97
2	SnTe$\langle I\rangle$ — 5.7 mol%	0.25	78	1.66
3	SnTe$\langle I\rangle$ — 22.4 mol%	0.25	59	1.85
4	SnTe$\langle I\rangle$ — 26.6 mol%	0.25	51	1.91
5	SnTe$\langle I\rangle$ — 14.1 mol%	0.25	68	1.75
6	Undoped SnO_2	0.35	50	2.11
7	SnTe$\langle I\rangle$ — 26.6 mol%	0.35	65	1.74
8	I — 5 at%	0.35	70	1.85

TABLE II. Ellipsometric angles for layers Nos. 6 and 7 on a quartz substrate at an angle of incidence of 55°.

No.	Ψ, deg	Δ, deg
6	18.95	164.64
7	9.52	206.50

obtained for layers prepared on quartz, Polykor, and ceramic substrates and also at different points in the same sample.

These layers were previously examined by atomic force microscopy to establish a correlation between the porosity, gas sensitivity, and refractive index.[7] An investigation of the surface topology showed that the gas-sensitive layers with a high impurity content have a substantially more developed surface relief. For instance, the characteristic size of the surface roughness at the base may differ by an order of magnitude, with typical dimensions of $0.1-1\ \mu$m.

The difference in the refractive indices of the layers determined by ellipsometry cannot be attributed merely to the surface roughness. The difference between the ellipsometric angles for layers of the same material with smooth and rough surfaces on substrates having the same surface structure should not exceed 2–3°, according to existing estimates,[8] which is substantially less than the difference obtained in the present investigation (Table II).

The correlation established between the gas sensitivity of the detectors and the refractive index of the sensor film suggests that the difference between the refractive indices of the tin dioxide layers reflects the different degree of porosity.

Although a porous material is an optically dissipative medium, no absorption was observed in the layers ($k=0$) in these ellipsometric investigations. Thus, the porosity of a material with a grain size of the order of the radiation wavelength is manifested as a reduction in the refractive index averaged over volume.

Table I and Fig. 1 reveal a good correlation between the gas sensitivity (characterized by the relative change in the resistance of the sensor $\Delta R/R$) and the effective refractive index, and the porosity of the film has a dominant influence on its gas sensitivity compared with the influence of well-known technological parameters such as the oxygen partial pressure. Figure 1 also gives data for undoped tin dioxide samples. The correlation between the gas sensitivity of these layers and their refractive index is satisfactorily described by the dependence shown approximately in the figure: the higher the porosity, the lower the refractive index of the SnO_{2-x} layer. We have therefore shown that laser interferometry can be considered to be an effective method for rapidly monitoring the porosity of layers to predict their gas sensitivity.

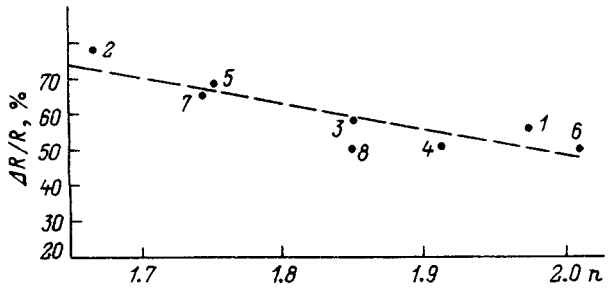

FIG. 1. Gas sensitivity of SnO_{2-x} layers for toluene ($C_{tol}=60$ mg/m^3) as a function of the effective refractive index.

The authors are grateful to A. Yu. Il'in for supplying data from studies of the microrelief of layers by atomic force microscopy.

This work was partially supported by Grant No. TsMID-37 from the Ministry of Education of the Russian Federation.

[1] S. R. Morrison, Sensors and Actuators, No. 12, 425 (1987).
[2] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light* [North Holland, Amsterdam, 1977; Mir, Moscow, 1981].
[3] T. Oyabu, Y. Ohta, and T. Kurobe, Sensors and Actuators, No. 9, 301 (1986).
[4] A. S. Bakin, M. V. Bestaev, D. Dimitrov *et al.*, Thin Solid Films **296**, 168 (1997).
[5] K. I. Dolotov, A. B. Zil'berman, Yu. L. Il'in *et al.*, Izv. Akad. Nauk Ross. Akad. Nauk Neorg. Mater., No. 1, 83 (1994).
[6] V. I. Pshenitsyn, M. I. Abaev, and N. Yu. Lyzlov, *Ellipsometry in Physicochemical Research* [in Russian], Khimiya, Leningrad (1986).
[7] M. V. Bestaev, D. Ts. Dimitrov, A. Yu. Il'in *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **62**, 549 (1998).
[8] G. V. Rozhnov, N. V. Bukhtiyarova, V. A. Kemarskiĭ *et al.*, in *Ellipsometry: Theory, Methods and Applications* [in Russian], Nauka, Novosibirsk (1991), pp. 34–44.

Translated by R. M. Durham

Efficiency of sequential resonance compression of radio pulses

S. N. Artemenko

Research Institute of Nuclear Physics, Tomsk Polytechnic University, 634050 Tomsk, Russia
(Submitted August 11, 1997; resubmitted April 21, 1998)
Zh. Tekh. Fiz. **69**, 131–133 (April 1999)

A study is made of the efficiency of forming high-power wide-band and ultrawide-band radio pulses by a sequence of resonance compression steps. An analysis is made of the compression of radio pulses with a rectangular envelope in a chain consisting of two or three compressors, and also in a chain where the output signal from the first stage has an exponential trailing edge and the other stages have rectangular envelopes. The maximum energy storage efficiency is determined for a cavity excited by a signal with an exponential trailing edge. © *1999 American Institute of Physics.* [S1063-7842(99)02704-X]

Novikov *et al.*[1] showed that high-power wide-band and ultrawide-band microwave pulses can be obtained by sequential resonance compression achieved by alternately storing and rapidly releasing the pulse energy in a linear chain of cavities. However, any increase in the number of compression stages clearly leads to additional energy losses. Consequently, the energy characteristics of these devices may be poor and depend strongly on the parameters of each stage. For instance, the efficiency of the device in Ref. 1 did not exceed 7% at a gain of ~ 28 dB. In this context the possibility of optimizing these systems to enhance the gain and the efficiency naturally arises.

In the present paper the gain and the efficiency are analyzed for the sequential compression of microwave pulses with a rectangular envelope and also for the compression of pulses with an exponential trailing edge.

1. Pulses having a nearly rectangular envelope are compressed in a chain of compressors with single-mode cavities. For this case the gain M^2 and the efficiency of energy transfer to the load $\varkappa$ for each stage are given by[2]

$$M^2=4\beta(1-\exp(-t_i(1+\beta)/2\tau_p))^2 \times(\tau_p/\tau_B-\beta-1)/(1+\beta)^2, \quad \varkappa=\eta_H\eta_p, \tag{1}$$

where β is the coupling coefficient between the cavity and the rf section, t_i is the input pulse duration, τ_p and τ_B are the cavity time constant and the attenuation constant of the output pulse, η_H is the energy storage efficiency, η_p is the cavity efficiency,

$$\eta_H=4\beta\tau_p(1-\exp(-t_i(1+\beta)/2\tau_p))^2/(t_i(1+\beta)^2) \quad \text{(from Ref. 3)},$$

$$\eta_p=1-Q_B/Q_H, \tag{2}$$

Q_H and Q_B are the cavity Q factors in the storage and release regimes.

Equations (1) and (2) readily yield a relationship between M^2, η_H, η_p, t_i, and $\varkappa$:

$$M^2=\eta_H\eta_p t_i/\tau_B=\varkappa t_i/\tau_B=\eta_H t_i(Q_0/Q_B-\beta-1)/\tau_B,$$

$$Q_0=Q_H(1+\beta). \tag{3}$$

For compressors with single-mode cavities the following relationships hold: $\tau_B\approx T$, $t_i>\tau_p$, $Q_B\ll Q_H$, so that Eqs. (1)–(3) are transformed to give

$$M^2=4\beta(1-\exp(-t_i(1+\beta)/2\tau_p))^2M_0^2/(1+\beta)^2,$$

$$\varkappa\approx\eta_H, \quad \eta_p\approx 1, \quad M^2\approx\eta_H t_i/T, \tag{4}$$

where $M_0^2=1/2\alpha$ is the cavity gain, α is the round-trip attenuation constant of the wave, and T is the round-trip time.

For a chain of N compressors we then obtain

$$M_N^2=t_1\prod_{i=1}^{N}\eta_i/T_N=t_1\varkappa_N/T_N, \quad \varkappa_N\approx\prod_{i=1}^{N}\eta_i, \tag{5}$$

where T_N is the round-trip time in the cavity in the last stage.

Estimates show that two or three stages are a realistic proposition. A larger number leads to an unjustified increase in the weight and size of the system and appreciably lowers the efficiency. The duration of the output signal from the first stage cannot exceed 30–40 ns, since any further increase results in an extremely long cavity at this stage, with a proportional decrease in the gain ($M_1^2\sim 1/T_1$). Thus, for the second and third stages the inequalities $t_{2,3}\ll\tau_{p2,p3}$, ($t_2=T_1$, $t_3=T_2$), are satisfied, which imply that the cavities at these stages must operate with strong overcoupling (β_2, $\beta_3\gg 1$). As a result we find $\eta_{2,3}\approx 0.8$ (Ref. 3), and for the two- and three-stage systems expressions (5) have the form

$$M_{N=2}^2\approx 0.8\eta_1 t_1/T_2, \quad M_{N=3}^2\approx 0.64\eta_1 t_1/T_3,$$

$$\varkappa_{N=2}\approx 0.8\eta_1, \quad \varkappa_{N=3}\approx 0.64\eta_1. \tag{6}$$

It can be seen from Eq. (6) that the gain of the system used to compress pulses with a rectangular envelope is a function of the storage efficiency of the first stage and the ratio of the durations of the input and output pulses, while the efficiency is only a function of the storage efficiency of the first stage. In the first stage fairly long cavities must be used, having an almost constant value of τ_{p1}, so that for given t_1 the optimum value β_1 is also constant; therefore, if the durations of the input and output pulses are fixed, the gain and efficiency of the chain will be constant almost regardless of the individual parameters of the first stage. Thus, in this particular case the choice of the first stage should be

dictated merely by the acceptability of its overall weight characteristics and electrical strength for a given input power level.

2. Single-mode cavities can deliver pulses with a rectangular envelope and have a high gain (up to 20 dB). However, because of their low Q factor and small volume, these cannot store a sufficiently large quantity of energy, and thus they cannot be used to obtain pulses with a high stored energy. Relatively high-Q multimode cavities are more suitable for this purpose, since these have a large volume with small overall dimensions and can form fairly long high-energy nanosecond microwave pulses ($\sim$100 ns). However, since they have a high storage efficiency ($\eta_1 \approx 0.6$), compressors with multimode cavities have a relatively low gain ($\sim$10 dB) and an output signal with an exponential trailing edge. This makes them uncertain prospects for use in the first stage, while their use in the second and third stages is very doubtful. Thus, we shall only consider a chain with a multimode cavity in the first stage and single-mode ones in the following stages. The output pulse envelope for a compressor with a multimode cavity is given by[2]

$$b_1^2(t) = 8\beta_1\beta_B(1-\exp(-t_1(1+\beta_1)/2\tau_p))^2 P_g \times \exp(-t/\tau_B)/(1+\beta_1)^2, \quad (7)$$

where β_B is the output coupling coefficient of the cavity, $\beta_B = \tau_p/\tau_B - \beta_1 - 1$, and P_g is the generator power.

It can be shown that in this case, the storage process in the second-stage cavity is described by the differential equation

$$db_2/dt + b_2/2\tau_{p2} \approx j\sqrt{\beta_2}\, b_1(t)/M_{01}T_2, \quad (8)$$

where M_{02} is the gain of the second-stage cavity.

From Eqs. (7) and (8) we find

$$b_2(t) = -4\sqrt{b_2}(\exp(-t/\tau_H) - \exp(-t/2\tau_B))\tau_H\tau_B b_1(0)/(T_2 M_{02}(2\tau_B - \tau_H)), \quad (9)$$

where $\tau_H = 2\tau_{p2}/(1+\beta_2)$.

The gain M_2^2 and storage efficiency η_2 of the second stage are then given by

$$M_2(t)^2 = 4M_{02}^2(\exp(-t/\tau_H) - \exp(-t/2\tau_B))^2/((1+\beta_2) \times (1-\tau_{p2}/\tau_B(1+\beta_2))^2, \quad (10)$$

$$\eta_2(t) = M_2(t)^2 T_2/\tau_B. \quad (11)$$

It can then be shown that $M_2(t)^2$ and $\eta_2(t)$ have a maximum at $t = t_m$, which is given by $t_m = 2z\tau_B \ln(z)/(z-1)$, where $z = \tau_{p2}/(\tau_B(1+\beta_2))$. Then $M_2(t)^2$ and $\eta_2(t)$ have the values

$$M_{2m}^2 = 4\beta_2\tau_B z(z^{(1/(1-z))} - z^{(z/(1-z))})^2/((1+\beta_2)(z-1)^2 T_2),$$

$$\eta_{2m} = M_{2m}^2 T_2/\tau_B. \quad (12)$$

These expressions yield an equation to determine the optimum input coupling β_2 of the second-stage cavity

$$(1-\beta_2-\tau_{p2}/\tau_B)/2\beta_2 - z\ln z/(1-z). \quad (13)$$

In general Eq. (13) can be solved numerically. However, in the limiting case when $\tau_{p2} \gg \tau_B$, it is easy to establish that $\beta_2 \approx \tau_{p2}/\tau_B - 1$. Then η_{2m} tends asymptotically to

$$\eta_{2m} = 4(1-\tau_B/\tau_{p2})/e^2 \approx 4/e^2 \to 0.54, \quad (14)$$

and the gain decreases monotonically as

$$M_{2m}^2 \approx 4M_{02}^2\tau_B/(e^2\tau_{p2}) \approx 4M_{02}^2/(e^2(\beta_2+1)) \to 4\tau_B/(e^2 T_2). \quad (15)$$

Thus, these last expressions indicate that when the cavity is excited by a signal with an exponential trailing edge, the storage efficiency cannot exceed 0.54 and the gain decreases monotonically as the exciting pulse becomes shorter, it tend to unity for $\tau_B \approx 2T_2$. However, since for oversize cavities one generally has $\tau_B > 10T_2 \approx 0.1\tau_{p2}$, the second-stage gain will be at least 5–10 in any case. For a chain of two or three compressors expressions (1), (10), and (11) readily yield expressions for M_N^2 and $\varkappa_N$:

$$M_{N=2}^2 = \varkappa_1\eta_2 t_1/T_2, \quad M_{N=3}^2 = 0.8\varkappa_1\eta_2 t_1/T_3,$$

$$\varkappa_{N=2} = \varkappa_1\eta_2, \quad \varkappa_{N=3} = 0.8\varkappa_1\eta_2, \quad (16)$$

where $\varkappa_1$ is obtained from Eq. (1) and η_2 is determined by formulas (12) and (13).

It follows from Eq. (16) that, unlike a chain of compressors with rectangular pulses, in a chain where the signal has an exponentially decaying edge the gain and efficiency after the first stage are determined by the energy transfer efficiency not only of the first stage but also of the second. Since generally $\varkappa_1 \approx 0.6$ (Ref. 2) and η_2 as given by Eq. (12) is 0.3–0.5, in this case the efficiency of the first two stages will be almost the same as that for a chain of single-mode cavities. This also applies to the gain. This confirms that a multimode storage cavity is undoubtedly promising for the first stage.

3. To conclude, this analysis has shown that for a sequential compression of pulses with a rectangular envelope after each compression stage and also of pulses with an exponential trailing edge after the first stage and a rectangular envelope in the following stages, the gain and efficiency of the system for fixed input and output pulse durations and a given number of stages are almost constant. This implies that there is some freedom in the choice of size of the first-stage cavities.

It has been established that the energy storage efficiency for microwave pulses with an exponential trailing edge does not exceed 0.54.

The author would like to thank Yu. G. Yushkov for supporting this work.

[1] S. A. Novikov, S. V. Razin, P. Yu. Chumerin, and Yu. G. Yushkov, Pis'ma Zh. Tekh. Fiz. **16**(20), 46 (1990) [Sov. Tech. Phys. Lett. **16**(10), 784 (1990)].
[2] S. N. Artemenko, Zh. Tekh. Fiz. **66**(10), 163 (1996) [Tech. Phys. **41**, 1058 (1996)].
[3] S. V. Baraev and O. P. Korovin, Zh. Tekh. Fiz. **50**, 2465 (1980) [Sov. Phys. Tech. Phys. **25**, 1444 (1980)].

Translated by R. M. Durham

Physical mechanisms for the development of lightning discharges between a thundercloud and the ionosphere

N. I. Petrov and G. N. Petrova

High-Voltage Research Center, 143500 Istra, Moscow District, Russia
(Submitted May 12, 1998)
Zh. Tekh. Fiz. **69**, 134–137 (April 1999)

An investigation is made of the influence of changes in atmospheric pressure with altitude and the thundercloud geometry on the development of lightning propagating upward to the ionosphere. It is shown that the mechanism for the development of high-altitude lightning does not differ from that for the formation and propagation of ordinary lightning between a thundercloud and the ground. It is established that high-altitude lightning forms as a result of a reduction in pressure with altitude and can only take place from thunderclouds located at high altitudes. © *1999 American Institute of Physics.* [S1063-7842(99)02804-4]

Studies of the physical mechanisms responsible for the development of discharges between thunderclouds and the ionosphere observed recently from artificial satellites and termed ''high-altitude lightning''[1,2] are important from the scientific viewpoint and to determine their influence on the characteristics of the ionosphere. Mechanisms usually considered responsible for the development of high-altitude lightning include the propagation of streamers in the quasielectrostatic field of a thundercloud[3,4] and breakdown by runaway electrons.[5,6] In the first case, the lightning is initiated and develops without any external factors being involved, whereas in the second case the initiation of the lightning is influenced by cosmic rays which generate high-energy electrons which, on entering the strong electric field of the thundercloud, are accelerated and create a flux of secondary avalanches. The second mechanism is supported by the fact that the electric fields in a thundercloud are an order of magnitude lower than the breakdown levels, and the initiation of discharges in the absence of some other factors cannot be explained. However, this mechanism requires experimental confirmation, whereas the first mechanism occurs in laboratory experiments. Extensive experimental data has been acquired to support the electrical nature of the formation and evolution of a leader discharge in long air gaps. However, laboratory experiments are usually carried out in nonuniform gaps with a particular electrode geometry and at near-normal pressures. It is therefore necessary to study how changes in atmospheric pressure with altitude influence the formation and evolution of high-altitude lightning. The influence of the thundercloud geometry on the formation of these discharges also requires investigation.

In the present paper we investigate how changes in atmospheric pressure with altitude and the thundercloud geometry influence the formation and evolution of lightning discharges from the top of thunderclouds to the ionosphere. We show that high-altitude lightning can be explained using existing concepts on the mechanisms for the development of streamer and leader discharges.

The development of high-altitude lightning is influenced by changes in pressure, atmospheric humidity, and thundercloud geometry. Studies of discharge processes as a function of these parameters are important to identify the physical mechanisms for the development of high-altitude lightning.

We shall take a dipole electric charge distribution as the thundercloud model. We know[7] that the upper part of the cloud is usually predominantly positively charged while the lower part is predominantly negatively charged, i.e., the thundercloud is an electric dipole. In addition to the main charges, small regions of positive charge may occur at the base of the thundercloud. The charged regions of the dipole have geometric dimensions of the order of a few kilometers.

The change in pressure with altitude is expressed by[8]

$$p(z)=p(0)\exp\left\{-\frac{\ln 10}{18400}z\right\}, \tag{1}$$

where $p(0)$ is the pressure at sea level, and z is the height above sea level in meters.

There exists a threshold field for breakdown, which is determined by the equality of the frequencies of ionization and attachment. For air this field is $E^*_{\rm dis}/p\approx 24\ {\rm V/(m\cdot Pa)}$. It should be noted that these fields are required to initiate a streamer discharge or an avalanche–streamer transition but not for streamer propagation. These fields may be found near microinhomogeneities (such as raindrops, charged particles, and so on) in a thundercloud. We know[9] that the development of a streamer discharge preceding the formation of a leader discharge depends on the air pressure, and the critical electric field strength required for the propagation of streamers decreases with decreasing pressure. Measurements in long air gaps show that at atmospheric pressure positive streamers propagate in a field $E^*_{\rm str}/p\approx 5\ {\rm V/(m\cdot Pa)}$ (Ref. 10). At an altitude of 20 km, where the pressure is $p=8.2\times 10^3$ Pa, the critical field for streamer propagation will be $E^*_{\rm str}\approx 41$ kV/m.

The condition for a streamer–leader transition must be satisfied to initiate a lightning discharge. This condition is usually satisfied if the electric field strength over the length of the streamer zone in the leader exceeds the critical value,

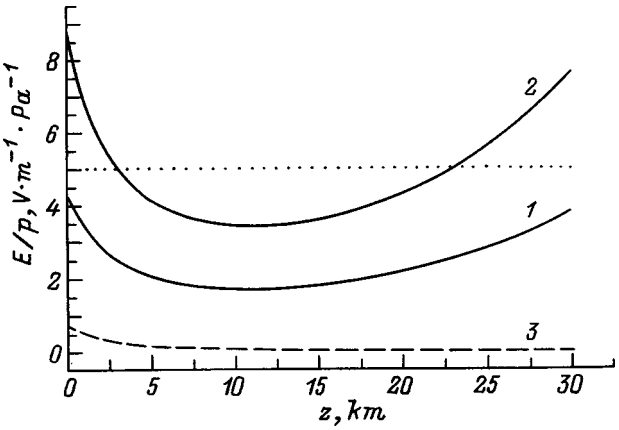

FIG. 1. Variation of E/p with distance z from a thundercloud of radius $R_0=5$ km for various cloud charges Q when the upper part of the thundercloud is at an altitude $h=20$ km above the Earth's surface: $p(z)=1$ atm = const; $Q=100$ (*1*) and 200 C (*2*, *3*).

$E \geqslant E^* = E^*_{\text{str}}$. Since the electric field strength decreases with distance from the thundercloud, the condition for a streamer–leader transition is only satisfied when the potential (charge) of the thundercloud exceeds the critical value. However, at high altitudes the condition for streamer propagation can be satisfied at substantially greater distances from the thundercloud because the pressure decreases exponentially with altitude. For a uniformly charged spherical thundercloud the electric field behavior may be expressed as follows:

$$E = \frac{\rho R_0^3}{3\varepsilon_0 r^2}, \quad r > R_0, \tag{2}$$

where R_0 is the radius of the thundercloud, ρ is the volume charge density, and $\varepsilon_0 = 8.85 \times 10^{-12}$ F/m is the permittivity of free space.

Measurements show that the volume charge density in thunderclouds reaches $\rho \approx 4$ nC, and the electric field strength is $|E_{\max}| \approx 50$–100 kV/m (Ref. 11).

Figure 1 gives $E(z)/p(z)$ as a function of the distance z from the upper edge of the thundercloud. It can be seen that the drop in pressure with altitude leads to a slower decrease in E/p with distance but that at high altitudes this quantity is even higher than it is near the thundercloud. This behavior of E/p as a function of altitude has the result that it is preferable for the discharge to propagate upward from the thundercloud rather than toward the ground. This is confirmed by the following results.

We shall use a method based on fractal theory[12,13] to simulate high-altitude lightning. In Ref. 12 we used this method to simulate the formation of lightning with allowance for intracloud discharges but we neglected the influence of changes in pressure with altitude. Here the fractal method is used to simulate lightning discharges for various values of the potential, charge, and altitude of the thundercloud above the ground neglecting and allowing for changes in atmospheric pressure with altitude. Figure 2 shows simulated patterns of breakdown trajectories from thunderclouds positioned at various altitudes. The calculations show that the

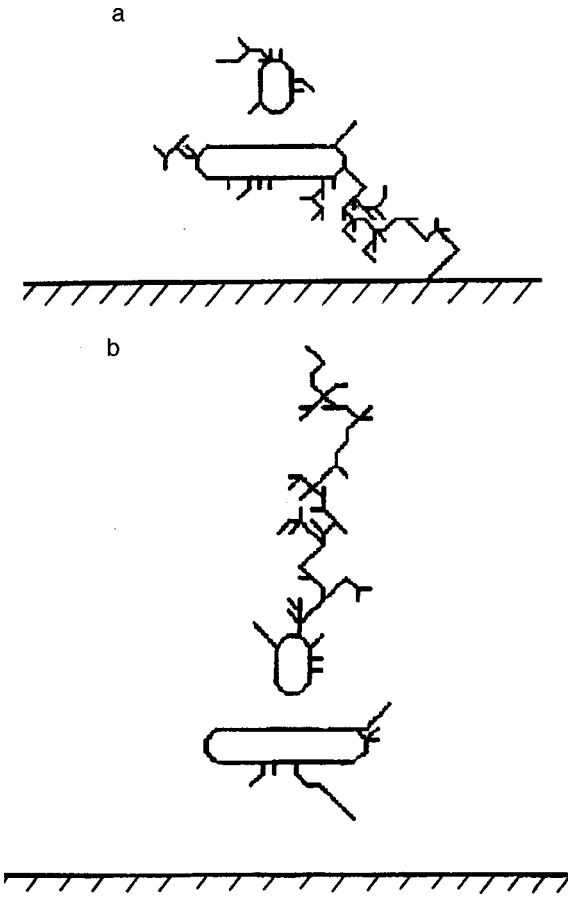

FIG. 2. Simulated patterns of breakdown trajectories, neglecting changes in atmospheric pressure with altitude (a) and allowing for pressure changes (b).

formation of upward-directed lightning is mainly influenced by changes in pressure with height above sea level. The development of upward lightning becomes preferred when the upper part of the thundercloud is situated at an altitude of around 20 km or higher. This is also confirmed by observations.[2] Since clouds are usually lower than this, high-altitude lightning is observed less frequently than cloud-to-ground discharges. For certain ratios of thundercloud parameters and altitude, discharges can develop both toward the ground and upward (Fig. 3).

The physical pattern of the development of high-altitude discharges is in many respects similar to the so-called incomplete discharges observed in long air gaps. The ''blue jet'' discharges correspond qualitatively to the development of the streamer zone of a positive leader, while the ''red sprites'' correspond to the processes in the streamer zone of a negative leader. For instance, the minimum propagation velocity of a positive streamer corona in air[14] and the propagation velocity of the luminous front of a blue jet discharge are the same ($v \approx 10^5$ m/s) and they also have the same geometric shape. We note that the streamer zone of a positive leader is a conical region filled with a branching structure of streamer channels at whose fronts the amplification of the field is sufficient for ionization. The same fine structure should be observed in blue jet discharges. In Refs. 3 and 4 blue jet discharges are represented as the development of a single streamer channel, but this model is incorrect, since the

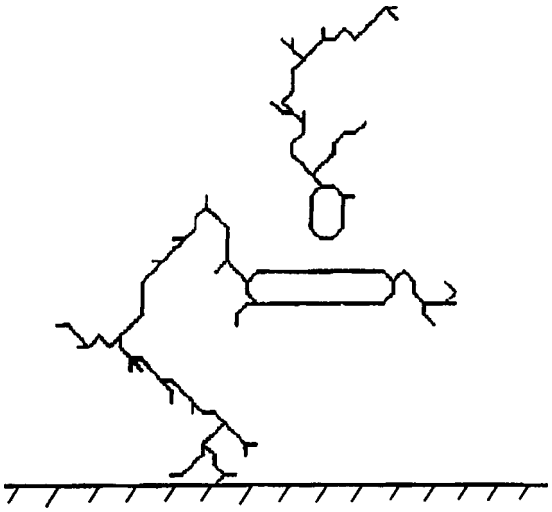

FIG. 3. Simulated pattern of a lightning trajectory propagating simultaneously toward the ground and upward.

field at the front of this streamer is inadequate for ionization.

Red sprite discharges generally follow positive cloud-to-ground lightning[2] and are observed at substantially higher altitudes compared with blue jets. The penetration of the quasielectrostatic field of a thundercloud is small at high altitudes because the medium is highly conducting. Thus, only a rapidly varying field penetrates at high altitudes, such as that created by discharges directed toward the ground. Qualitatively, red sprite discharges are very similar to discharges from plasmoids formed at the front of the streamer zone in a negative leader in laboratory experiments in the atmosphere. The mechanism for the formation of these plasmoids is presently unknown. They may be formed as a result of the action of the electromagnetic field of high-power lightning. However, we can postulate that the nature of the formation of plasmoids in laboratory experiments and in high-altitude lightning is the same. One of the most probable mechanisms may be the focusing of streamers (plasma filaments) in a nonuniform electric field. This is because streamer formations are polarized in an external electric field, i.e., they acquire dipole moments. Particles in a thundercloud also acquire dipole moments and this promotes their focusing in regions of strong field. Estimates made assuming that the Coulomb repulsion forces are equal to the forces of attraction of the dipole moments show that this mechanism does not contradict the observations. This same mechanism may result in the formation of cells in a thundercloud. We know that a thundercloud has a cellular structure. Cells with different densities of excess charge were also observed when cumulus clouds growing into thunderclouds were probed by airplane.[15] Lightning is clearly initiated in these zones of excess charge density.

It is interesting to determine the parameters of high-altitude discharges and compare them with the parameters of discharges in laboratory conditions. For instance, using the Gauss theorem we can estimate the charge introduced by the streamer zone of a leader from the known electric field strength at the boundary of the streamer zone E_{str} as

$$Q_{\mathrm{str}}=\varepsilon_0 \oint E\,dS \approx 2\pi\varepsilon_0 E_{\mathrm{str}} l_{\mathrm{str}}^2\left(1-\cos\frac{\Omega}{2}\right), \qquad (3)$$

where Ω is the apex angle at the tip of the streamer zone, l_{str} is the length of the streamer zone, and $\varepsilon_0=8.85\times10^{-12}$ F/m is the permittivity of free space.

The electric field at the front of a positive streamer zone at an altitude of 40 km is $E_{\mathrm{str}}^{*}\approx3.4$ kV/m. The apex angle of the streamer zone is usually $\Omega\approx15°$ (Ref. 4) and the length of the streamer zone is $l\approx20$ km. Substituting these values into formula (3) we find that the charge in the streamer zone is $Q\approx0.7$ C which gives an average electron concentration $n_e\approx10^7\ \mathrm{m}^{-3}$. In laboratory discharges the charge of the streamer zone is usually $Q\sim10^{-5}$ C. Estimates made from direct measurements of the luminous exitance of high-altitude discharges give $n_e\approx10^7\ \mathrm{m}^{-3}$, i.e., they agree with our results. The intensity of the discharge optical radiation is determined by the densities of the air molecules and by the rates of excitation of radiative transitions[16] and may be expressed in terms of the charge concentration n_e as follows:[4]

$$I\approx2\times10^{-2}\frac{n_e}{(1+3\times10^{-4}\cdot n_e^{1/2})},\ R.$$

Substituting the average value for the charge concentration $n_e\approx10^{-7}\ \mathrm{m}^{-3}$, we obtain the radiation intensity $I\approx10^5$ R ($1R=10^{-10}$ photon$\cdot$m$^{-2}\cdot$s^{-1}) which also agrees with the direct measurements.

To conclude, the development of high-altitude lightning is strongly influenced by changes in pressure with altitude and for a given size of thundercloud there are critical potentials at which a discharge may form. This explains why lightning propagating from the tops of clouds to the ionosphere is only observed at high altitudes above 20 km where the pressure of the atmosphere is less than one-tenth of the pressure at sea level.

This work was supported financially by the Russian Fund for Fundamental Research (Grant No. 97-02-16382-a).

[1] W. L. Boeck *et al.*, Geophys. Res. Lett. **19**, 99 (1992).
[2] W. L. Boek, J. Geophys. Res. **100**, 1465 (1995).
[3] V. P. Pas'ko, U. S. Inan, and T. F. Bell, Geophys. Res. Lett. **23**, 301 (1996).
[4] A. I. Sukhorukov, E. V. Mishin, P. Stubbe, and M. J. Rycroft, Geophys. Res. Lett. **23**, 1625 (1996).
[5] A. V. Gurevich, G. M. Milikh, and R. Roussel-Dupre, Phys. Lett. A **165**, 463 (1992).
[6] A. V. Ivanovskiĭ, Pis'ma Zh. Tekh. Fiz. **22**(12), 6 (1996) [Tech. Phys. Lett. **22**(6), 480 (1996)].
[7] M. A. Uman and E. P. Krider, IEEE Trans. Electromagn. Compat. **EMC-24**(2), 79 (1982).
[8] *Handbook of Physical Quantities*, edited by I. S. Grigor'ev and E. Z. Meĭlikhov [in Russian], Energoatomizdat, Moscow (1991).
[9] C. T. Phelps and R. F. Griffiths, J. Appl. Phys. **47**, 2929 (1976).
[10] N. I. Petrov, V. R. Avanskiĭ, and N. V. Bombenkova, Zh. Tekh. Fiz. **64**(6), 50 (1994) [Tech. Phys. **39**, 546 (1994)].
[11] W. D. Rust *et al.*, in *Proceedings of the Ninth International Conference on Atmospheric Electricity*, St. Petersburg, Vol. 1, 160 (1992).

[12] N. I. Petrov and G. N. Petrova, Zh. Tekh. Fiz. **63**(4), 41 (1993) [Tech. Phys. **38**, 287 (1993)].
[13] N. I. Petrov and G. N. Petrova, *Abstracts of Ga URSI*, Lille, France, 1996, p. 230.
[14] N. I. Petrov, V. R. Avanskiĭ, and N. V. Bombenkova, Pis'ma Zh. Tekh. Fiz. **19**(23), 34 (1993) [Tech. Phys. Lett. **19**(12), 751 (1993)].
[15] I. M. Imyanitov, *Electrification of Aircraft in Clouds and Precipitation* [in Russian], Gidrometeoizdat, Leningrad (1970), 210 pp.
[16] J. Chamberlain, *Physics of the Aurora and Airglow* (Academic Press, New York, 1961).

Translated by R. M. Durham

Generation and focusing of shock–acoustic waves in a liquid by a multicenter electric discharge

V. S. Teslenko, A. I. Zhukov, V. V. Mitrofanov, and A. P. Drozhzhin

M. A. Lavrent'ev Institute of Hydrodynamics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted May 15, 1998)
Zh. Tekh. Fiz. **69**, 138–140 (April 1999)

Results are presented from an investigation of the generation and focusing of shock–acoustic waves by a multicenter discharge in an electrolyte in the shape of a segment of a sphere of radius $R=170$ mm and aperture $D=145$ mm. It is shown that compared with the well-known electric-discharge generators of shock–acoustic waves, the development of this type of generator holds great promise. © *1999 American Institute of Physics.* [S1063-7842(99)02904-9]

At present, shock–acoustic wave generators are finding increasing applications in medicine, underwater acoustics, and in scientific research. Topical problems involve forming shock–acoustic waves (~ 0.1–1 μs) of a particular geometry in a given space. The following methods of generating and focusing shock–acoustic waves are currently available for noninvasive lithotrypsy, surgery, and therapy:[1,2] I — using point electrical breakdown in water focused by a reflector in the shape of a truncated cavity of an ellipsoid of revolution; II — using an electromagnetic emitter, where a concave metal membrane is set in motion by electromagnetic radiation; III — using piezoelectric elements mounted in a mosaic pattern on a segment of a sphere or other profile, IV — using laser methods with various geometric profiles.

In the present paper we consider the possibility of generating and focusing shock–acoustic waves created by a multicenter discharge in a liquid.[3] This is a discharge where numerous plasma centers are formed above the surface of one of the electrodes, outside the interelectrode gap and open to the shock–acoustic wave radiation. The shape of this surface may be specified. Here we report the first results of electrodynamic investigations for a multicenter electric-discharge emitter in the form of a segment of a sphere of radius $R=170$ mm with a radiation working aperture $D=145$ mm.

The apparatus is shown schematically in Fig. 1. A positive emitting electrode *1* in the form of a segment of a sphere and having peaking devices *2* was prepared by special technology as a periodic grating structure in an insulating frame-

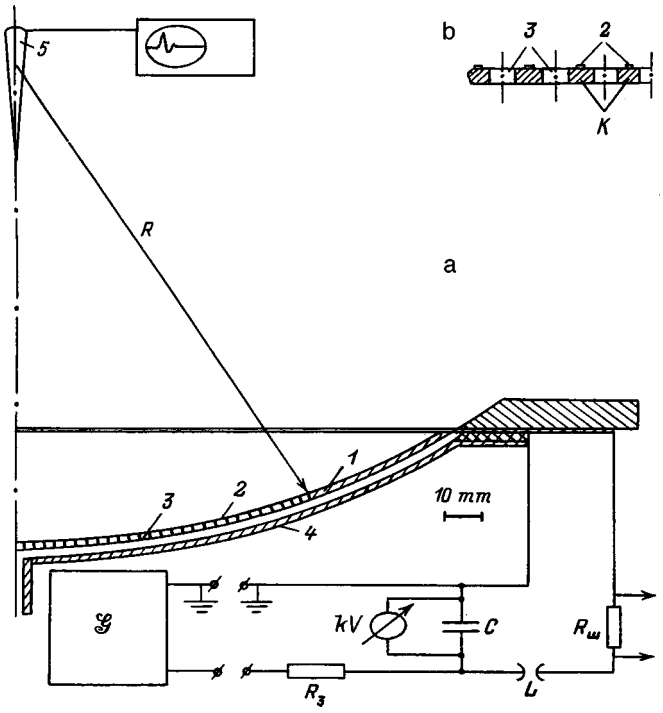

FIG. 1. Schematic of multicenter electric-discharge generator of shock–acoustic waves with focusing (a) and magnified multicenter electrode (b).

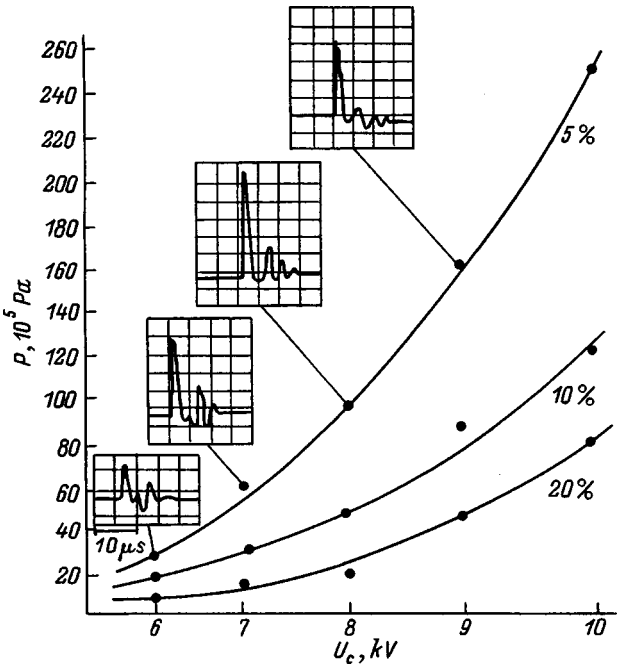

FIG. 2. Results of measurements of the peak parameters of shock–acousti c waves at the focus for three electrolyte concentrations as a function of the capacitor voltage.

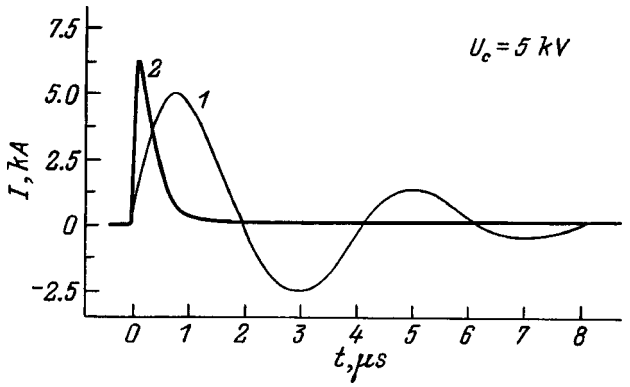

FIG. 3. Oscilloscope traces of the current for geometrically similar shock–acoustic wave generators using the same current source: *1* — electromagnetic, *2* — multicenter.

work K with apertures *3* ($d=1$ mm). The entire surface of the electrode, except for the peaking devices, was covered with insulator. The steel peaking devices had an average end area $s \approx 0.15\ \text{mm}^2$ and height $h \approx 0.1$ mm. The distance between the peaking devices was $l \approx 3$ mm, and the number of these emitting peaking devices was around 2000. A copper negative electrode *4* comprising a spherical segment of diameter $D=220$ mm was positioned at a distance $H=4.5$ (± 0.5) mm below the positive electrode plate. The electrode system was immersed in an electrolyte solution. A capacitance $C=2\ \mu$F charged from a current source (Y) was discharged across the working electrodes *1* and *4* by means of a controllable spark gap L. The pressure field was measured using a tourmaline piezoelectric transducer *5* with 0.5 mm spatial resolution and 0.05 μs temporal resolution. The discharge current was monitored using a low-inductance shunt $R_s = 10^{-3}\ \Omega$ and an S-9-27 oscilloscope connected to a personal computer. Multicenter discharges were generated in an aqueous solution of 5–20% sodium chloride at voltages between 1 and 10 kV.

Figure 2 gives results of measuring the peak parameters of the parameters of the shock–acoustic waves at the focus for three concentrations of the electrolyte solution (5%, 10%, and 20% NaCl) for capacitor voltages up to 10 kV.

We established that an increase in the concentration of the electrolyte solution reduces the amplitude of the emitted shock–acoustic wave. This indicates that the eletroacoustic efficiency depends on the electrolyte concentration. At low discharge energies the oscilloscope traces of the pressure pulses show a rarefaction wave. As the discharge energy increases, the rarefaction wave may be converted into a compression wave because of cavitation processes in the focal region.[4]

A comparison between this multicenter shock–acoustic wave generator and a geometrically similar electromagnetic generator ($R=170$ mm, nine turns, sector working diameter $D=220$ mm, gap between planar coil and membrane 250 μm) for the same parameters of the initial current source and energy store C shows that the multicenter generator has advantages over well-known electric-discharge generators because the discharge time can be reduced. Figure 3 shows oscilloscope traces of the current for geometrically similar electromagnetic (trace *1*) and multicenter generators (trace *2*) with the same capacitive energy store. These traces show that by replacing the inductive load with an active one, we can reduce the discharge time and therefore the duration of the shock–acoustic wave. We note that this electromagnetic generator has parameters close to the optimum for this type of generator.[5]

We know that the sharpness of focusing and the amplitude of the focused shock–acoustic wave depend on the parameter $k=D/F$ (F is the focal length), i.e., the larger k, the better the focusing of the wave. However, for electromagnetic generators an increase in the parameter k (for fixed F) increases the inductance L ($L \sim D$) which therefore increases the discharge time $\tau \sim R\,L\,C$ and the duration of the shock–acoustic wave. Reducing the inductance of electromagnetic generators by reducing the gap between the coil and the membrane is limited by the electrical strength of the insulating spacers. For a multicenter shock–acoustic wave generator (with the same energy store C), the situation is the reverse. In this case, if the electrical conductivity γ of the electrolyte is fixed, an increase in the aperture leads to an increase in the number of peaking devices n and the total electrode area $S=n\,s$, which reduces the load resistance $R \approx H/(\gamma S)$, the discharge time, and hence the duration of the generated and focused shock–acoustic waves. The larger number of peaking devices means better erosion resistance compared to type I point electrical-breakdown generators. We also note that for voltages above 8 kV the discharge time may decrease nonlinearly as a result of electrodynamic processes at the peaking devices, but this topic is outside the scope of the present article.

This work was supported by the Russian Fund for Fundamental Research Grant No. 96-02-19329.

[1] Kh. Raikhenberg, Radiotekh. Elektron. **78**(9), 194 (1988).
[2] Yu. V. Andriyanov, A. A. Li, and V. S. Teslenko, Vopr. Kurortol. Fizioterapii Lechebnoĭ Fizkul't. No. 3, 54 (1993).
[3] V. S. Teslenko, A. I. Zhukov, and V. V. Mitrofanov, Pis'ma Zh. Tekh. Fiz. **21**(18), 20 (1995) [Tech. Phys. Lett. **21**(9), 733 (1995)].
[4] V. S. Teslenko, Pis'ma Zh. Tekh. Fiz. **20**(5), 51 (1994) [Tech. Phys. Lett. **20**(3), 199 (1994)].
[5] Yu. V. Andriyanov and S. P. Kryuchkov, *Acoustics of Inhomogeneous Media (Dynamics of Continuous Media 110)* [in Russian], Novosibirsk (1995), pp. 12–18.

Translated by R. M. Durham

Myxine design for an experimental Galathea reactor

A. I. Morozov, V. A. Nevrovskiĭ, V. I. Pistunovich, and A. N. Svechkopal

Kurchatov Institute Russian Science Center, 123182 Moscow, Russia
(Submitted March 2, 1998)
Zh. Tekh. Fiz. **69**, 141–142 (April 1999)

[S1063-7842(99)03004-4]

Galathea plasma traps[1] contain current-carrying conductors immersed in the plasma. These conductors are called myxines and are specific to the Galathea,[2] so that analyses of possible myxine designs are urgently required for the development of a Galathea fusion reactor.

In the present paper we consider fundamental aspects of the design of a myxine with a superconducting chord for a demonstration reactor, i.e., a Galathea with a pulse duration of around 1000 s. The myxine parameters should be maintained during burning of a DT plasma, from which a neutron flux of 1 MW/m^2 reaches the myxine surface. The main features of the myxine design described below were reported at conferences.[3,4]

We consider a three-zone (or triple-shell) radiation-accumulative myxine in which the outer (''red'') shell is heated to $T\sim 2000$ K and effectively radiates up to 90% of the incident energy flux from the plasma (1 MW/m^2). The next (''gray'') layer is separated from the red by a multilayer heat-insulated vacuum shield and should reduce the energy flux to the superconductor by a factor of approximately 1000. Finally, the third (''blue'') zone is the cryozone. A superconducting Nb_3Sn winding is immersed in hydrogen slush (a mixture of liquid and solid hydrogen). Here the temperature of the superconductor is determined by the melting point of hydrogen $T=14$ K. In the working state the temperature may rise to 17 K (evaporation point of hydrogen at 1 atm), at which Nb_3Sn is still a superconductor. The cryozone is also separated from the gray zone by a heat-insulated vacuum shield to reduce the heat flux to the cryozone. Our task is to determine the transport and absorption of neutrons and gamma rays in this myxine design. We report results of preliminary calculations of the energy distribution over the minor diameter of the myxine, which is taken to be 1.7 m with its volume divided into 15 layers. The calculations were made using the MCNP code[5] recommended for neutronics calculations at ITER. The specific heat of the materials which can be used in the myxine varies between 1.6 and 3.6 J/cm^3 K for C, Si, W, and Fe, respectively. Estimates show that the average myxine temperature would reach ~ 2000 K after a 1000 s fusion pulse if there is no cooling. Thus, radiation of most of the energy by the red layer limits the myxine temperature. We shall analyze each myxine zone in greater detail.

The main constraints on the red zone are that it should absorb approximately 90% of the neutron and gamma-ray energy, the materials should have a low vapor pressure at 2000 K, and the thickness of this zone should be as small as possible. Bearing in mind these constraints, we selected tungsten, which can reduce the neutron and gamma energy flux tenfold at a low vapor pressure ($\sim 10^{-11}$ Torr at ~ 2000 K) with a thickness of 10–15 cm. Its thermal conductivity is ~ 100 W/mK and the thermal diffusivity is ~ 0.4 cm^2/s. The most important characteristic of tungsten is its large neutron albedo, which in our model reached $\sim 60\%$.

The thickness of the gray zone with the heat-insulating vacuum shield was selected as 45 cm. The rate of temperature rise of the layer was estimated here neglecting the thermal conductivity of the shield ($\lambda\sim 0$). This zone is mainly stainless steel. In addition, three water-based layers used to moderate the fast neutrons and reduce the energy release in the gray zone.

The blue zone (cryozone) contains a Nb_3Sn superconducting chord immersed in a hydrogen slush with a heat of fusion of 4.2 J/cm^3. The minor radius of the cryozone is 25 cm, the cooled volume of one linear meter of the myxine is approximately 0.2 m^3/m, the thickness of the superconducting coil is around 5 cm, with a minor radius of 20 cm, and the total cross section of the superconductor is ~ 630 cm^2. The permissible energy at the superconductor is ~ 0.8 MJ/m. The power flux to the Nb_3Sn is limited to 800 W/m for a 1000 s pulse.

This myxine design presupposes that there are thermal bridges passing through the heat-insulating vacuum shield and linking the gray zone with the red zone and the cryozone. Estimates show that the heat fluxes flowing across the thermal bridges are considerably lower than the incoming energy flux with the neutrons and gamma rays. We propose that tungsten should be used in the red zone for structural strength and that stainless steel should be used in the gray and blue zones. The myxine cross section is shown in Fig. 1. The thickness of the red zone (tungsten) is 15 cm, the thickness of the gray zone with the heat-insulating vacuum shield is 45 cm, and the cryozone radius is 25 cm.

The heat distribution over the myxine thickness from the neutrons and gamma rays was calculated for several sets of materials. The simplest version is shown in Table I, which indicates that 15 cm of tungsten is sufficient to absorb most of the energy in the red zone, and the release in the cryozone is then reduced approximately 1000 times.

We then determine the current in the myxine required to form a magnetic layer 50 cm thick around the myxine, between the plasma boundary and the myxine surface, to prevent fast alpha particles from reaching the myxine. For a

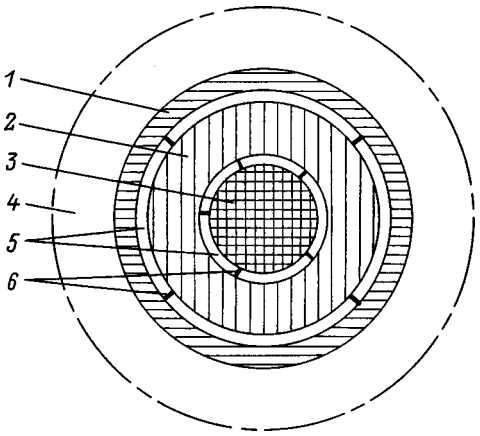

FIG. 1. Cross section of myxine: *1*—red zone, *2*—gray zone, *3*—blue zone, *4*—magnetic layer of myxine, *5*—vacuum shields with thermal insulation *6*.

magnetic field $B \sim 1$ T at the outer boundary of the magnetic layer, the current in the superconductor should be $I_m \sim 7$ MA for the chosen dimensions. Estimates show that the total mass per linear meter of myxine together with the structure is around 15 tonnes. In this case, levitation of the myxine may be sustained by an additional transverse magnetic field of around 0.03 T.

The myxine is one of the principal components of the Galathea so that an estimate of its cost will give some idea of the overall cost of a Galathea fusion reactor. At present it is difficult to make an absolute cost estimate. However, its cost may be compared with that of the ITER experimental tokamak fusion reactor.

The design of the myxine may be compared with one section of the ITER fusion reactor, having a single toroidal magnetic coil together with a radiation shield, i.e., 1/20 of the reactor. From this it follows that the cost of the myxine should not exceed 1/20 of the cost of the magnetic system with the radiation shield in the ITER reactor.

TABLE I. Distribution of neutron and gamma-ray energy release over fifteen layers of myxine cross section.

Layer	Material	W, W / cm^3
1	Hydrogen	$1.82E-03$
2	Nb_3Sn	$8.39E-03$
3	Hydrogen	$2.11E-03$
4	Ice	$3.97E-03$
5	SS	$1.56E-02$
6	SS	$3.02E-02$
7	SS	$1.06E-01$
8	Ice	$4.52E-02$
9	SS	$1.67E-01$
10	SS	$2.63E-01$
11	SS	$8.38E-01$
12	Water	$4.30E-01$
13	Tungsten	$2.08E+00$
14	Tungsten	$2.35E+00$
15	Tungsten	$5.38E+00$

Note: Each layer is 5 cm, layers 1–3 form the cryozone, layers 4–12 form the gray zone, and layers 13–15 form the red zone; layers 4, 8, and 12 are water, and layers 1 and 3 are hydrogen slush.

[1] A. I. Morozov, Pis'ma Zh. Tekh. Fiz. **16**(15), 86 (1990) [Sov. Tech. Phys. Lett. **16**(8), 597 (1990)].
[2] A. I. Morozov and V. I. Khripunov, Fiz. Plazmy. **18**, 838 (1992) [Sov. J. Plasma Phys. **18**, 435 (1992)].
[3] A. I. Morozov, V. A. Nevrovsky, and V. I. Pistunovich, in *Proceedings of the Sixth Conference on Engineering Fusion Problems*, St. Petersburg, 1997.
[4] A. I. Morozov, V. A. Nevrovsky, and V. I. Pistunovich, IAEA Technical Committee meeting on ''Innovative Approaches to Fusion,'' Pleasanton (USA), 1997.
[5] *MCNP—A General Monte Carlo Code for Neutron and Photon Transport*, Version 3a, edited by J. Briesmeister, Report No. LA-7396-M, Rev. 2, Los Alamos National Laboratory (1986).

Translated by R. M. Durham